ȘTEFAN ȘTERIE

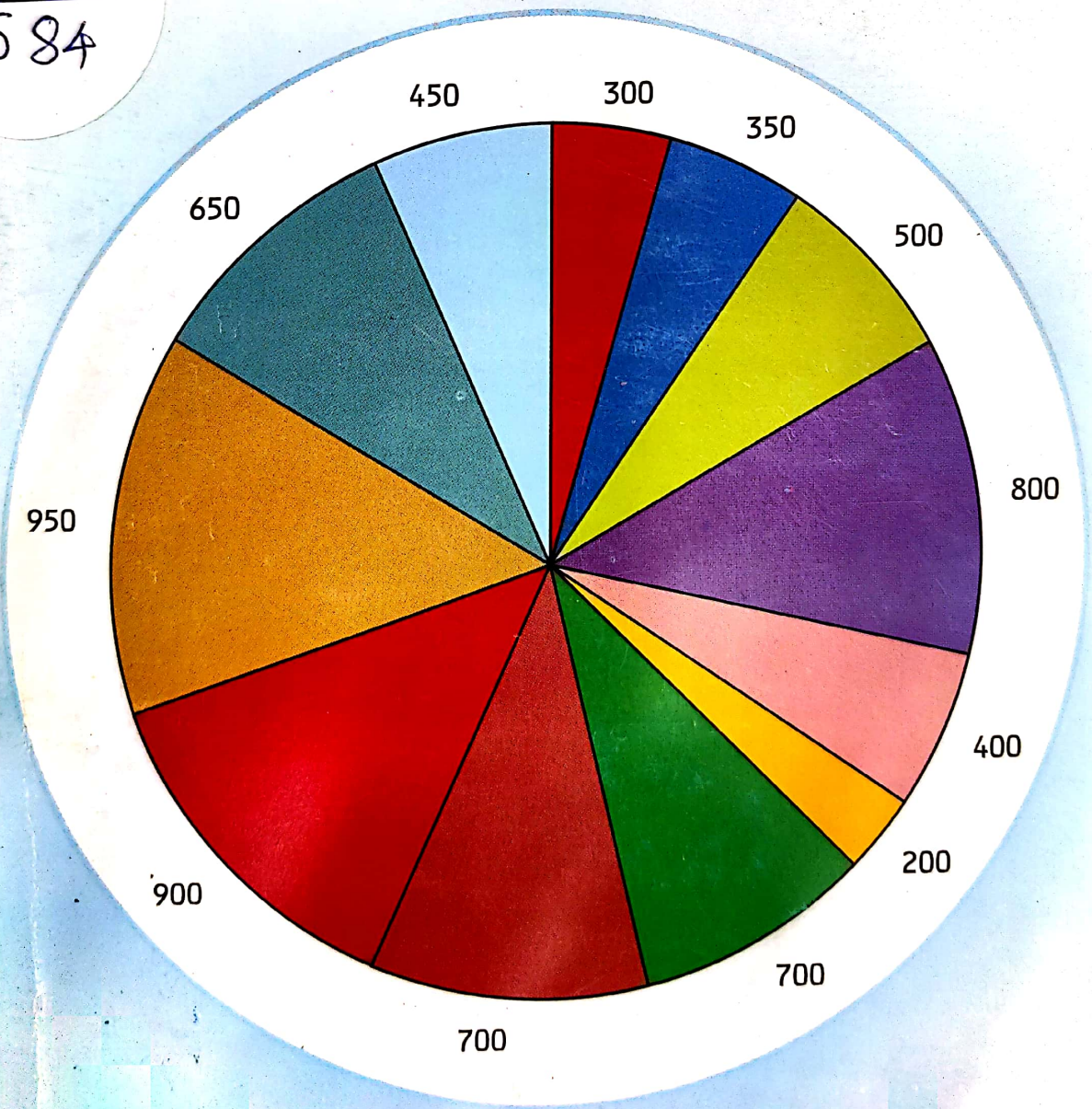
TITI PARASCHIV

AMADO ȘTEFAN

STATISTICĂ

TEORIE - APLICAȚII - PROGRAME

311
S84



Scanned with OKEN Scanner

ŞTEFAN STERIE

TITI PARASCHIV

AMADO ŞTEFAN

STATISTICĂ

TEORIE — APLICAȚII — PROGRAME

Scanned with OKEN Scanner

Prefață

Procedeele statistice sunt folosite pentru construirea și verificarea ipotezelor de lucru utile în fundamentarea deciziilor. Nevoia de statistică are la bază necesitatea pentru evaluarea și utilizarea informației despre evenimente și fenomene economice, fizice, sociale, medicale, naturale. Fiecare aspect al activității umane este măsurat și interpretat în termeni statistici, adică măsurat sub formă numerică și interpretat în medie. Cuprinși de puterea numerelor, trăim senzația că numai cunoscându-le putem stăpâni fenomenele și evenimentele.

Statistica reține, prin procedee specifice, ceea ce este esențial și logic în vederea obținerii indicatorilor statistici și descoperirii regularităților, legităților și tendințelor în evoluția fenomenelor.

Statistica teoretică are ca scop definirea fundamentelor și funcțiilor gnoseologice, a sistemului general de indicatori statistici, a metodelor de observare, de grupare și de centralizare, de calcul al valorilor medii, al dispersiei, corelației, seriilor dinamice, indicatorilor selecției, precum și de analiză, adaptare la specificul fenomenelor social economice de masă.

Conceptele de bază ale statisticii matematice sunt frecvența unui eveniment sau a unei caracteristici într-o populație, legea de repartiție sau de distribuție a unei mărimi, variabilitatea sau stabilitatea unei mărimi, corelațiile sau conexiunile între două sau mai multe caracteristici, indicii statistici.

Conceptele de câmp de probabilitate sau statistic, de valoare medie, de abatere, medie pătratică, de variabilă statistică sau aleatoare, de funcție de repartiție, de independență sau dependență, de corelație sau de regresie, legate de procesele aleatoare, prelucrate de teoria probabilităților servesc ca elemente fundamentale ale statisticii matematice și edificării teoriilor ei specifice. Printre aceste teorii specifice figurează teoria selecției și a estimației, a testelor de semnificație, teoria deciziilor și teoria predicției.

Prezenta lucrare se adresează în special studenților din facultățile de științe economice, dar poate fi utilă și acelora care sunt interesați în cunoașterea fenomenelor și proceselor care se află sub semnul variabilității și incertitudinii.

Autorii

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 1

INTRODUCERE ÎN STATISTICĂ

1.1. Chestiuni preliminarii

Statistica este o ştiinţă complexă cu un arsenal diversificat de metode şi cu o arie vastă de aplicaţii, fiind rezultatul unei îndelungate evoluţii. Ea a apărut din necesităţi practice, şi anume din nevoia reală de a cunoaşte în expunere numerică diferitele fenomene din natură şi societate.

Statisticizarea este un proces care a cuprins în orbita sa toate ştiinţele datorită necesităţii prelucrării volumului din ce în ce mai mare de rezultate.

Statistica este ştiinţa care descrie şi analizează numeric, cantitativ determinările calitative ale fenomenelor de masă, supuse acţiunii legilor statistice, dezvăluind particularităţile lor de volum, structură, conexiune, dinamică, precum şi legile care le guvernează. Sub raport gnoseologic, statistica are următoarele faze:

a) Observarea — culegerea informaţiei despre procesul sau obiectul luat în studiu;

b) Formularea ipotezelor;

c) Predicţia — deducţiile ce se fac pe baza ipotezelor formulate;

d) Validarea — este etapa de generalizare şi verificare a predicţiilor formulate prin culegerea de noi rezultate sau efectuarea de experimente.

Multitudinea de definiţii date statisticii are determinări *istorice, metodologice* şi *aplicative*.

Determinările istorice privesc faptul că statistica s-a constituit ca o ştiinţă interdisciplinară cu un profund caracter social.

Sub raport *metodologic*, statistica include un sistem de metode în special cantitative, cu fundamentarea epistemologică specifică.

Prin generalitatea metodelor sale, statistica are *aplicații* într-o paletă largă, de la astronomie la societate, de la micro-cosmos la macro-cosmos, de la fizica statistică la statistica socială.

Statistica s-a născut din necesități practice, evoluția sa către știință realizându-se în următoarele etape:

a) *Statistica descriptivă*
b) *Statistica analitică sau funcțională*
   — aritmetică;
   — probabilistică;
   — informațională;
   — sistemică.

În prezent, statistica dispune de un arsenal de metode și de o teorie generală.

Cultura statistică devine tot mai mult o componentă a culturii generale, iar gândirea statistică o manieră de analiză și interpretare a fenomenelor.

Apărută din nevoia de a surprinde cantitativ fenomenele social-economice, statistica a apărut în mediul social, extinzându-și aria de investigație și particularizându-se ca știință în nenumărate domenii, cum ar fi:

— statistica matematică;
— fizica statistică;
— statistica termodinamică;
— mecanica statistică;
— statistica economică;
— statistica juridică.

Teoria statistică sau statistica definește sistematic conceptele, noțiunile, principiile și metodele de cercetare statistică.

Sub raport metodologic, statistica are două componente distincte:

- *statistica descriptivă* reprezintă componenta clasică a statisticii, care are ca obiect de studiu descrierea statistică;
- *statistica inductivă* reprezintă componenta care analizează prin metode specifice datele experimentale; ea mai poartă denumirea și de *inferența statistică*.

Scanned with OKEN Scanner

Din perspectivă economică, statistica se poate organiza în trei categorii:

— statistica descriptivă;
— statistica matematică;
— statistica economică.

1.2. **Istoricul statisticii**

Există consemnări istorice precum că operații statistice (evidența populației, a terenurilor, a grânelor) s-au efectuat încă din anii 2300 î.Ch., în China; egiptenii efectuau evidențe ale terenurilor, populației şi recoltelor în mod regulat încă din anii 1900 î.Ch. În textele biblice, în Grecia antică sunt menționate înregistrări pentru ca în imperiul roman să capete un aspect regulat — recensăminte (*census*). În timpul împăratului Augustus s-a făcut bilanțul bogățiilor imperiale, pentru ca în timpul lui Antonius Pius să devină obligatorie declararea naşterilor, înființându-se şi birouri în care se păstrau registrele fiscale, vamale, cadastrale etc.

Investigațiile au fost diversificate în Evul Mediu:

— inventarierea de bunuri ale feudelor, mânăstirilor, breslelor şi corporațiilor;
— însemnările naşterilor, căsătoriilor, deceselor la parohii;
— lucrări de evidență din inițiativa statului.

Contabilizarea resurselor în antichitate şi Evul Mediu a avut scopuri pragmatice, efectuându-se fără criterii riguroase.

Depăşirea stadiului de *statistică practică sau materială* în exclusivitate s-a produs în momentul în care, peste interesele practice imediate, s-au evidențiat interese de ordin ştiințific, de analiză cu mijloace matematice a fenomenelor consemnate.

a. *Etapa descriptivă* (universitară). În secolele al XIII-lea şi al XIV-lea, statistica reprezenta descrierea statului — expunerea situației geografice, economice, politice — perfecționându-se în perioada Renaşterii; preluată de la italieni, descrierea statului a atins apogeul în secolele al XVII-lea şi al XVIII-lea în Germania, unde a căpătat un caracter mult mai organizat prin Şcoala descriptivă germană fondată de Herman Conrig (1600÷1681). Gottfried Achenwall (1719÷1772)

Scanned with OKEN Scanner

continuă învățământul statistic la Universitatea Göttingen, unde descrierea statului devine disciplină de predare academică.

Etimologic, cuvântul *statistică* provine de la latinescul medieval *status*, „stat" sau „stare politică" prin intermediul substantivului italian *statista*, cu înțelesul de „om politic versat în treburile statului".

A. L. Schlözer (1735÷1809) a exprimat obiectul statisticii în trei cuvinte:

Vires unitae agunt, care au următoarea semnificație:

vires se traduce prin forțele naturale și dobândite (oamenii, teritoriul și avantajele naturale);

unitae – forma și organele de guvernământ;

agunt – guvernarea și administrația.

Statistica este istorie în repaus, istoria – statistică în mișcare.

W. Hooper definește statistica astfel:

„Știința despre organizarea politică a statelor moderne ale lumii".

Büsching dă statisticii o formă numerică remarcabilă utilizând evidențele oficiale.

Statistica a intrat în istoria disciplinelor științifice ca o disciplină descriptivă, orientată spre informație și nu spre căutarea de legi.

b. *Faza aritmeticii politice.* În timp ce în universități statistica se configura ca disciplină descriptivă a statului, în afara universităților din Anglia se năștea o statistică total deosebită sub numele de aritmetică politică, ceea ce presupune:

— analiza datelor de observație prin procedee matematice;

— desprinderea regularităților în fenomenele social-economice;

— formularea de previziuni.

Folosirea metodei analitice, preconizată de filosofii materialiști metafizicieni ai timpului, recurgerea la instrumentul matematic și căutarea legităților prefigurează statistica modernă.

Reprezentanții școlii aritmetice politice din Anglia au pus accentul mai ales pe cercetarea fenomenelor demografice, deoarece vedeau în populație izvorul bogăției și puterii acestei țări.

John Grannt (1620÷1674) pune în evidență legitățile creșterii populației, ale echilibrului numeric dintre sexe, ale fertilității, mortalității, realizând chiar o tabelă de mortalitate.

Scanned with OKEN Scanner

William Petty (1623÷1689), întemeietorul economiei politice moderne, este creatorul denumirii *Aritmetica politică;* în cartea cu acelaşi nume face un studiu al fenomenelor social-economice în expresie cantitativă. Utilizează date şi calcule numerice în compararea potenţialului uman şi economic al diferitelor state şi, în particular, în evaluarea posibilităţilor comerciale şi maritime ale Angliei în rivalitatea sa cu Franţa.

Şcoala aritmeticii politice engleze este preluată de germani, care prin Johann Peter Süssmilch (1707÷1767) realizează o sinteză a aritmeticii politice prin problematicile abordate:

— căutarea explicaţiei fenomenelor vieţii umane şi indicarea legilor ce le guvernează;
— corelarea fenomenelor demografice şi economice;
— examinarea fenomenelor în dinamică;
— identificarea regularităţilor în masa de observaţii;
— alegerea şi evaluarea critică a datelor.

Aritmetica politică s-a dezvoltat foarte mult în demografie, construindu-se tabele de mortalitate în mai toate ţările occidentale, tabele care au constituit punctul de plecare al industriei de asigurări.

c. *Faza probabilistă.* În timp ce descrierile statistice devin obiect al instituţiilor publice, domeniul conceptual al aritmeticii politice capătă o nouă dimensiune odată cu introducerea calculului probabilităţilor.

Primele încercări teoretice au fost făcute de Jacob Bernoulli, care a formulat „legea numerelor mari", a deosebit probabilitatea *apriori* (matematică, logică sau deductivă) de probabilitatea *aposteriori* (statistică, empirică sau inductivă) şi a schiţat posibilitatea aplicării calculului probabilităţilor în economie.

Pierre-Simon Laplace (1749÷1827) a evidenţiat avantajele în studiul fenomenelor dependente de cauze complexe şi a demonstrat utilizarea posibilă a acestui calcul în exploatarea rezultatelor statistice privind fenomenele sociale.

Karl Friederich Gauss (1775÷1855) defineşte legea normală de repartiţie.

Simon Denis Poisson (1781÷1840) descoperă legea de repartiţie a evenimentelor rare studiind probleme de demografie, de statistică morală şi judiciară, cu aplicaţii ale calculului probabilităţilor.

Scanned with OKEN Scanner

Adolph Quételet (1796÷1874) înlocuieşte noţiunea de *homo economicus* al clasicilor cu *homo statisticus*, construind un fel de fizică socială conform căreia caracterele fizice, intelectuale şi morale ale oamenilor, observaţi în masă, s-ar echilibra într-o fiinţă fictivă, ideală, dorită de natură în jurul căreia oamenii reali s-ar distribui la întâmplare, analog erorilor în măsurarea unei mărimi.

Conceptul de om mediu implică noi concepte, cum ar fi: *repartiţie, medie, dispersie, observare de masă, regularitate,* esenţiale în cercetările statistice.

Concepută ca o fizică socială, statistica lui Quételet încorporează ideea că regularităţile descoperite prin observarea de masă rezultă din legi care se impun omului în ordinea fizică şi morală. Statistica devine o teorie sintetică, cu un arsenal de principii, tehnici şi procedee fiind denumită „anatomia societăţii"; ea are drept obiectiv observarea modificărilor pe care le suferă popoarele în starea lor fizică şi morală.

Aria de investigaţie a statisticii îşi lărgeşte problematica prin cercetarea variaţiei şi a relaţiilor cauzale în masa fenomenelor, dezvăluirea regularităţilor, investigarea cu predilecţie a fenomenelor vieţii umane.

Interpretarea probabilistă a statisticii este continuată de A. Cournot (1801÷1877), după care statistica este ştiinţa care are ca obiect culegerea şi coordonarea faptelor de orice fel, încât să se obţină raporturi numerice sensibil independente de anomaliile hazardului şi care denotă existenţa cauzelor regulate a căror acţiune s-a combinat cu cea a cauzelor întâmplătoare.

Începuturile statisticii româneşti sunt marcate de două personalităţi proeminente: Ion Ionescu de la Brad (1818÷1891) şi Dionisie Pop Marţian (1829÷1865).

Statistica modernă. Sfârşitul secolului XIX marchează o nouă etapă în evoluţia procesului de conceptualizare a statisticii, şi anume etapa constituirii statisticii moderne, în cadrul căreia gândirea s-a deplasat spre interpretarea analitică a fenomenelor de masă şi obţinerea de concluzii inductive pe baza observaţiilor empirice. Această schimbare de conţinut a statisticii a determinat întrepătrunderea cu matematica, în special prin şcoala anglo-saxonă de statistică matematică ai cărei

Scanned with OKEN Scanner

principali reprezentanți sunt Fancis Galton (1822÷1911), Karl Pearson (1857÷1936) și Ronald Aylmer Fisher (1890÷1962).

„Variația legată" a variabilelor observate ale colectivității statistice, dezvăluită de Galton, transpusă apoi în limbaj matematic de Pearson, se constituie într-o teorie coerentă denumită *teoria corelației statistice.*

R. A. Fisher conturează obiectul statisticii ca fiind alcătuit din trei categorii de probleme:

- probleme de specificație — expresia matematică a legii de repartiție a populației;
- probleme de estimație — depistarea metodelor convenabile pentru calculul coeficienților proprii estimației parametrilor necunoscuți ai populației;
- probleme de repartiție — repartiția estimațiilor parametrice și a unor coeficienți aleși pentru a testa validitatea specificației.

În domeniul modern conceput de Fisher, se pornește de la ipoteze statistice asupra procesului sau fenomenului observat, consecințele lor logic deduse se compară cu datele disponibile, stabilindu-se concordanța cu ipotezele stabilite.

În accepțiunea modernă, statistica este preocupată de generalizarea rezultatelor obținute pe eșantion la întregul sistem, înlocuindu-se *observarea totală* cu *sondajul statistic*, estimându-se totodată gradul de incertitudine al concluziilor inductive.

În diversitatea definițiilor, se disting două opinii:

— statistica — metodă de cercetare;

— statistica — disciplină științifică cvasi-independentă.

d. *Etapa informațională* a gândirii statistice începe cu teoria matematică a comunicației a lui Claude Shannon (1948).

Teoria informației este o ramură a matematicii, adiacentă calculului probabilităților și ciberneticii și care studiază problemele matematice ale obținerii, transmiterii, stocării și transformării informației. Modelul lui Shannon este statistico-probabilistic și conține drept noțiune fundamentală entropia informațională *H*.

$$H = -\sum p_i \log p_i \qquad (1.1)$$

unde: p_i reprezintă probabilitățile sau frecvențele semnalelor sursei.

Scanned with OKEN Scanner

e. *Etapa sistemică* a statisticii reprezintă un sistem de noțiuni, organizat în jurul conceptului de probabilitate, de metode științifice, elaborate de statistica matematică, cu fundamentare epistemologică explicită și cu o varietate de domenii de aplicare, acoperind practic natura și societatea.

Cercetările științifice simt necesitatea unei „escorte statistice" în două momente fundamentale ale demersului cognitiv:

— momentul empiric — culegerea și prelucrarea faptelor brute;

— momentul aplicativ — verificarea teoriilor pe calea confruntării cu realitatea.

Statistica pune la dispoziție tehnici și procedee adecvate operațiilor întreprinse în fiecare din cele două momente metodologice ale cunoașterii științifice.

Pe terenul ori în vecinătatea statisticii au apărut numeroase discipline și teorii noi, cum sunt „metriile" (biometria, econometria, psihometria, sociometria, tehnometria, jurimetria), cercetările operaționale, cibernetica, teoria informației, teoria deciziilor.

În sensul cel mai larg, statistica este matricea oricărei științe experimentale, iar ca furnizor principal de informații este indispensabilă activității de conducere.

Teoria probabilităților este teoria matematică ce se ocupă cu studiul fenomenelor întâmplătoare sau aleatoare utilizând în acest scop metodele deductive ale matematicii. Fenomenelor aleatoare de masă le este caracteristică proprietatea de stabilitate a frecvențelor. Conceptul fundamental de probabilitate corespunde intuitiv constantei în jurul căreia oscilează frecvențele și către care șirul acestor frecvențe tinde când numărul probelor crește nelimitat (Legea numerelor mari a lui J. Bernoulli).

În teoria probabilităților există două definiții:

— definiția clasică a probabilității;

— definiția axiomatică a probabilității.

Probabilitatea caracterizează gradul de probabilitate a producerii unui eveniment în condiții bine determinate.

Statistica matematică este disciplina matematică consacrată elaborării noțiunilor și metodelor specifice studiului statistic al fenomenelor de masă. Într-o formulare generală, statistica matematică se ocupă cu

Scanned with OKEN Scanner

gruparea, analiza şi interpretarea datelor referitoare la anumite fenomene, precum şi unele previziuni privind producerea lor în viitor.

Statistica matematică include :

– *teoria sondajului i*— formularea principiilor şi procedeelor de prelevare a eşantioanelor şi de prelucrare a datelor de observaţie potrivit cu regulile calculului probabilităţilor, în scopul de a dobândi cantitatea de informaţie necesară estimării parametrilor populaţiei reprezentate ;

– *teoria estimaţiei* — construirea formulelor care permit evaluarea parametrilor unei populaţii pe baza informaţiei rezultate din cercetarea pe eşantion ;

– *verificarea ipotezelor statistice* — elaborarea metodelor cu ajutorul cărora se judecă valabilitatea ipotezelor privind parametrii sau forma de repartiţie a unei populaţii plecând de la eşantion ;

– *teoria deciziilor statistice* — stabilirea metodelor de inferenţă statistică aplicabile la luarea deciziilor în condiţii de incertitudine.

Statistica economică este disciplina care are ca obiect studiul statistic al fenomenelor şi proceselor economice de masă.

Ea are două componente distincte :

— statistica economică generală ;

— statistica economică de ramură.

Capitolele importante ale statisticii economice generale sunt :

— statistica resurselor de muncă ;

— statistica avuţiei naţionale ;

— statistica producţiei materiale, etc.

Domeniul predilect al statisticii economice este metoda balanţelor, seriile cronologice, indicii, analiza tendinţelor.

Scanned with OKEN Scanner

1.3. **Obiectul şi metoda statisticii**

Definirea şi delimitarea obiectului statisticii presupune determinarea *ariei de investigaţie, a particularităţilor obiectului* şi a *raportului dintre statistică şi celelalte ştiinţe care studiază fenomene şi procese ale domeniilor conexe.*

a) **Aria de investigaţie**

Statistica studiază regularităţile fenomenelor de masă sau a fenomenelor de tip colectiv.

Fenomenele de masă se produc sub acţiunea unor factori cu caracter sistematic, asociaţi cu cei întâmplători în cadrul unor relaţii de cauzalitate dinamice.

În funcţie de aria de investigaţie, statistica are două componente distincte:

— o componentă ce studiază fenomenele şi procesele din natură;

— o componentă ce studiază fenomenele din societate — statistica social-economică.

Fenomenele *simple* sau *certe* sunt fenomenele în care relaţia cauză-efect este imediată şi univocă adică unei cauze îi corespunde un efect şi numai unul.

Există şi fenomene care apar ca rezultat al combinării complexe a unei mulţimi dinamice de factori şi influenţe care determină apariţia unor fenomene cu multiple manifestări individuale aparent întâmplătoare.

Determinarea trăsăturilor specifice unei colectivităţi trebuie cercetată colectivitatea în totalitatea ei. Fenomenele caracteristice unei colectivităţi se numesc *fenomene de masă* şi ele reprezintă obiectul de studiu al statisticii.

Analiza statistică este de tip cantitativ, dar ea reprezintă cuantificarea trăsăturilor calitative ale fenomenelor de masă. Principala proprietate a acestor fenomene este variabilitatea în timp şi în spaţiu, iar legea de apariţiei a acestora se manifestă ca tendinţă ce poate fi cunoscută doar la nivelul ansamblului.

Particularităţile statisticii social-economice sunt :

— obiectul său de studiu îl reprezintă fenomenele social-economice de masă ;

Scanned with OKEN Scanner

— caracterizează cantitativ-numeric fenomenele social-economice.

Metodologia de investigaţie statistică a fenomenelor porneşte, utilizând instrumentele formale furnizate de matematică, de la trăsăturile caracteristice ale fenomenelor de masă, de aceea putem delimita statistica matematică de statistica domeniului de investigare.

Metodologia, statistica matematică operează *deductiv*, pe când statistica domeniului are drept cale tipică metoda inductivă.

Metodologia statistică reprezintă totalitatea operaţiilor, tehnicilor, procedeelor şi metodelor de investigare statistică a fenomenelor ce aparţin unor procese de tip stohastic.

Sub raport metodologic, statistica reprezintă instrument de investigaţie pentru caracterizarea cantitativ-numerică efectuată de celelalte ştiinţe care studiază societatea ca sistem sau componente ale acestuia.

Etapele studiului statistic sunt:

a) ***Observarea*** care are ca finalitate cunoaşterea manifestărilor cantitative individuale.

Formele de realizare a observării sunt :

— recensăminte ;
— anchete ;
— monografii ;
— sondaje.

b) ***Prelucrarea rezultatelor observării*** care are ca obiectiv determinarea trăsăturilor specifice esenţiale ale evoluţiei fenomenului observat.

Metodele de prelucrare specifice statisticii sunt :

— metoda grupărilor ;
— metoda mediilor ;
— metoda analizei variaţionale şi de corelaţie ;
— metoda indicilor.

Rezultatele prelucrării se reflectă în indicatorii statistici.

Analiza şi interpretarea rezultatelor şi formularea concluziilor statistice.

Spre deosebire de statistica matematică, statistica socială are, sub raport metodologic, următoarele particularităţi :

— studiile sale au întotdeauna un caracter concret, bazându-se pe cunoaşterea empirică;

— variabilitatea fenomenelor sociale are amplitudine mai mare datorită combinării multitudinii de factori cauzali.

Scanned with OKEN Scanner

Complexitatea fenomenelor de masă care formează obiectul statisticii a impus separarea statisticii pe *ramuri :*

— Teoria statisticii sau Bazele statisticii ;
— Statistica macroeconomică ;
— Statistica descriptivă ;
— Statistica de ramură ;
— Statistica teritorială ;
— Statistica mediului înconjurător.

Apariția cibernetícii, informaticii, teoriei sistemelor a lărgit aria de investigație a statisticii, de aceea ea poate fi considerată o disciplină de graniță care prin metodele sale asigură cercetărilor științifice un caracter interdisciplinar.

Statistica oficială este compusă din :

— Instituții de statistică ale statului sau internaționale;
— Activitatea instituțiilor de statistică în domeniul culegerii, prelucrării şi publicării datelor statistice.

Instituțiile de statistică pot fi :

— oficii centrale de statistică ;
— birouri naționale de statistică ;
— institute centrale de statistică ;
— direcții centrale de statistică.

Cele mai importante atribuții ale instituțiilor de statistică se realizează prin :

— efectuarea de recensăminte ;
— efectuarea de anchete statistice ;
— înregistrarea rezultatelor economice ;
— întocmirea conturilor naționale ;
— publicarea de anuare statistice.

Gestionarea sistemului informațional statistic din țara noastră revine C.N.S. (Comisei Naționale pentru Statistică), organism care are ca principal obiectiv caracterizarea statistică a fenomenelor şi proceselor economice şi sociale la nivel macro şi microeconomic.

Obiectivele C.N.S. se reflectă în organizarea acesteia pe *domenii de activitate* şi *teritorial*. Publicațiile oficiale ce conțin rezultatele activității C.N.S. sunt :

Buletin de informare ;
Buletin lunar de prețuri ;

Scanned with OKEN Scanner

Bugete de familie ;
Anuarul statistic al României ;
Studii în reviste de specialitate ;
Comunicările datelor statistice la organismele internaţionale.

Dintre organismele internaţionale care au şi atribuţii în domeniul statistic sunt :

— O.N.U. în cadrul căruia funcţionează Comisia de statistică a O.N.U., iar în cadrul Secretariatului O.N.U., biroul de statistică ;
— Consiliul Economic şi Social (ECOSOC), în cadrul căruia funcţionează Comisia de statistică ;
— Organizaţia pentru cooperare şi dezvoltare economică (OCDE), în cadrul căreia funcţionează Departamentul Economic şi Statistic ;
— Comunitatea Europeană care are Oficiu de statistică ;
— Institutul inter-american de statistică ;
— Institutul internaţional de statistică ;
— Comisii economice regionale.

Manifestări ştiinţifice :

— congrese internaţionale de statistică (primul în 1853) ;
— conferinţe ale statisticienilor europeni.

Ca o concluzie a celor discutate până aici, putem da definiţia statisticii.

Statistica este ştiinţa care studiază aspectele cantitative ale determinărilor calitative ale fenomenelor de masă, fenomene care sunt supuse acţiunii legilor statistice care se manifestă în condiţii concrete, variabile în timp şi spaţiu.

1.4. **Categoriile ştiinţei statistice**

a. ***Fenomen de masă (fenomen colectiv, fenomen statistic)***

Fenomenul exprimă trăsăturile individuale, exterioare, întâmplătoare şi vremelnice ale obiectelor, trăsături prin care domeniile realităţii se manifestă nemijlocit şi multiplu.

Fenomenul de masă se constituie sub acţiunea colectivă şi repetată a unui număr mare de factori întâmplători; de aceea, punerea în evidenţă a esenţei reclamă studiul unui număr mare de cazuri.

Scanned with OKEN Scanner

Fenomenul de masă reprezintă rezultatul raportului dintre necesitate şi întâmplare, dintre legea statistică şi legea dinamică, dintre modelul stohastic şi cel determinist.

Fenomenul de masă este rezultat al lanţurilor de evenimente cauzale în care sunt integrate scopurile umane — individuale şi sociale.

b. ***Lege statistică (stohastică, lege de probabilitate)*** reprezintă legea ce guvernează comportamentul fenomenelor de masă, descriind relaţii cauzale statistice ; ea se realizează ca o tendinţă predominantă.

Legea statistică este una din formele de manifestare a conexiunii universale a fenomenelor din natură şi societate.

Noţiunea opusă şi complementară este legea dinamică (deterministă) conform căreia, valorilor strict determinate ale unor factori le corespund întotdeauna valori strict determinate ale mărimilor dependente de aceşti factori ; raporturile cantitative dintre mărimi rămân adevărate pentru fiecare caz individual. Exemplu : legea atracţiei universale.

Legea dinamică descrie comportamentul riguros al sistemelor sumative şi închise a căror evoluţie poate fi determinată ; legea statistică exprimă comportamentul sistemelor complexe, integrative şi deschise.

c. ***Relaţii cauzale statistice*** (legături statistice) reprezintă relaţii prin care se realizează mecanismul procesului de determinare şi dezvoltare a fenomenelor de masă.

Relaţiile cauzale mecanice sunt descrise de legi dinamice, relaţiile cauzale statistice sunt descrise de legi statistice şi exprimă relaţii neunivoce, relaţii în care cauzele interacţionează cu factorii aleatori într-un cerc larg de condiţii, astfel încât trecerea de la cauză la efect este mediată de jocul întâmplărilor, cuprins într-un model statistic (stohastic).

În succesiunea genetică a fenomenelor funcţie de ansamblul condiţiilor care însoţesc dinamica lor obiectivă, legea cauzalităţii se manifestă diferit funcţie de varietatea formelor de mişcare şi a nivelurilor de organizare a materiei.

În structurarea relaţiilor cauzale neliniare care au loc în procesele aleatoare stohastice, precum şi în cercetarea mecanismului lor de funcţionare intervine probabilitatea.

d. ***Legea numerelor mari*** (J. Bernoulli, 1713). Teoremă a statisticii matematice potrivit căreia frecvenţa unui eveniment converge în probabilitate către probabilitatea acelui eveniment.

Scanned with OKEN Scanner

Fie ∂_n numărul aparițiilor unui eveniment A în n experiențe independente și cu $p=p(A)$ probabilitatea evenimentului A, atunci :

$$\lim_{n\to\infty} P\left(\left|\frac{\partial_n}{n} - p\right| < \varepsilon\right) = 1 \quad (\forall) \qquad \varepsilon > 0 \qquad (1.2)$$

Altfel spus : La un număr mare de experimente abaterile individuale de la tendința centrală se pot compensa, sau altfel :

$$\text{frecventa}\,\frac{\partial_n}{n} \to p \qquad \text{dacă} \quad n \to \infty \ . \qquad (1.3)$$

Când mecanismul aleator, care permite să se atribuie *apriori* o valoare evenimentului, este cunoscut, frecvența constatată empiric pe un mare număr de cazuri va fi luată drept măsură a probabilității evenimentului.

e) ***Model (statistic)***

Modelul este o verigă într-un proces epistemologic mai larg care integrează :

— ipoteza ;
— reprezentarea schematică a unui proces prin model ;
— testarea statistică a modelului ;
— reluarea procesului, în cadrul unei teorii generale.

Modelul reprezintă, în general, transcrierea matematică a unor ipoteze simplificate despre starea sau evoluția unui fenomen sub influența anumitor variabile. După tipul variabilelor, se disting două modele :

— *Modelul determinist* este modelul care nu conține variabile aleatoare și care determină cert starea sau evoluția fenomenului.

— *Modelul statistic (stohastic)* este modelul în care intervin, pe lângă variabilele măsurabile sau observabile, una sau mai multe variabile aleatoare.

Variabilă endogenă (variabilă dependentă) este variabila ale cărei valori sunt deduse din rezolvarea modelului.

Variabilă exogenă (variabilă independentă) este variabila introdusă din afară, ale cărei valori sunt independente de model.

Variabila dependentă este funcție de variabila independentă.

f. ***Populație (statistică)*** reprezintă o mulțime de elemente care au cel puțin o caracteristică (calitate) comună și care face obiectul

Scanned with OKEN Scanner

cercetării statistice. Pentru ca rezultatele observării să fie consistente, populația (statistică) trebuie să fie delimitată în :

— conținut ;
— spațiu ;
— timp.

g. ***Eşantion statistic*** reprezintă submulțimea sau partea dintr-o populație statistică, constituită după anumite criterii, pentru a servi la determinarea caracteristicilor populației.

h. ***Unitate statistică*** reprezintă un element al unei populații statistice, asupra căruia se va efectua nemijlocit observarea ; ea reprezintă purtător originar de informație sau, altfel spus, subiectul logic al informației statistice.

i. ***Caracteristica statistică*** reprezintă însuşirea comună unităților unei populații, reținută în tematica investigației statistice şi care capătă accepții sau valori diferite de la o unitate la alta sau de la un grup de unități la altul.

j. ***Variabilă statistică*** reprezintă variația unei caracteristici. Ea poate fi calitativă sau cantitativă. Variabila cantitativă poate fi continuă sau discretă.

k. ***Variabilă aleatoare***. Categorie fundamentală a teoriei probabilităților desemnând variabila ale cărei valori, *apriori* necunoscute, apar în împrejurări datorate întâmplării, cu probabilități determinate.

Variabilele aleatoare pot fi : — continue ;
— discrete.

Fie variabila aleatoare discretă :

$$X\begin{pmatrix} x_1, & x_2, \dots & x_n \\ f(x_1), & f(x_2), \dots f(x_n) \end{pmatrix}, \qquad (1.4)$$

în care :

$$f(x_i) = P(X = x_i); \qquad \sum_{i=1}^{n} f(x_o) = f(x_1) + f(x_2) + \dots + f(x_n) = 1$$

l. ***Repartiție statistică***

Fie o variabilă statistică X care ia valorile $x_1, x_2, \dots x_n$ cu frecvențele $f_1, f_2, \dots, f_n$. Mulțimea ale cărei elemente sunt perechile ordonate (x_i, f_i) constituie repartiția variabilei statistice X:

Scanned with OKEN Scanner

$$X:\begin{pmatrix} x_1, x_2, \ldots, x_n \\ f_1, f_2, \ldots, f_n \end{pmatrix} \qquad (1.5)$$

m. ***Parametru statistic (valoare tipică)*** este valoarea reprezentativă dedusă dintr-o operație numerică aplicată la o repartiție statistică. Ea este rezultatul celei mai mari condensări posibile a informațiilor furnizate de o masă de observații.

Nivelul repartiției reprezintă ordinul de mărime caracterizat de valoarea centrală a variabilei în jurul valorii centrale.

Formă a repartiției reprezintă îndepărtarea repartiției statistice de la o repartiție teoretică.

Parametrii statistici se clasifică astfel :

- *Parametrii de nivel* : media, mediana, modulul ;
- *Parametrii de dispersie* : varianta, abaterea medie pătratică, coeficientul de variație ;
- *Parametrii de asimetrie și de boltire (aplatizare)* : coeficienții lui Pearson, coeficienții lui Fisher.

Tot în categoria de parametri mai sunt cuprinși :

— indicii de concentrare ;
— coeficienții de corelație ;
— media de sondaj.

n. ***Indicator statistic.*** Mărime statistică ce caracterizează un fenomen social-economic sub raportul structurii, interdependențelor, al modificării sale în timp sau în spațiu.

După expresia matematică, indicatorii statistici pot fi :

— mărimi absolute ;
— mărimi derivate.

Scanned with OKEN Scanner

CAPITOLUL 2

SERII STATISTICE. INDICATORI NUMERICI

Statistica studiază aspectele cantitative ale fenomenelor variabile în timp şi spaţiu, pornind de la conţinutul calitativ al acestora.

Statistica este o disciplină de graniţă cu aplicaţii în: economie, fizică, biologie, medicină, matematică, sociologie, etc.

Statistica poate fi: — *descriptivă;*
— *analitică (funcţională).*

Statistica descriptivă se ocupă cu ordonarea faptelor (fenomenelor) pe baza observării lor exterioare. Principala ei operaţie este clasificarea.

Statistica descriptivă **operează cu valori medii** şi abateri, dar este insuficientă pentru înţelegerea structurală şi dinamică a evenimentelor.

Apoi intervine statistica analitică (statistică matematică) care grupează datele, le analizează şi le interpretează în vederea unei predicţii privind comportarea viitoare a fenomenului.

Noţiuni utilizate în statistica matematică

Statistica operează cu noţiunile de :

— populaţie statistică ;
— efectivul total al populaţiei statistice ;
— caracteristică ;
— frecvenţă.

Definiţia 1.

Numim **populaţie statistică** orice mulţime care formează obiectul unei analize statistice. Elementele unei populaţii statistice se numesc unităţi statistice (indivizi, subiecţi).

Definiţia 2.

Trăsătura comună tuturor unităţilor unei populaţii, care ne interesează în cadrul unei analize statistice, se numeşte ***caracteristică***.

O caracteristică se numeşte cantitativă dacă se poate măsura. În caz contrar caracteristica se numeşte calitativă.

Scanned with OKEN Scanner

Caracteristicile cantitative pot fi :

— **discrete** (discontinue) — cu valori numerice numere naturale
(nr. Firmelor dintr-un oraş, judeţ),
(nr. locuitorilor dintr-un oraş).

— **continue** (caracteristica poate la orice valoare într-un interval finit sau infinit). (cazul taliei, greutăţii; lungimi, volume, cifre de afaceri, producţie fizică, producţie marfă...)

Exemple de caracteristici calitative :

- culoarea părului, ochilor unor persoane;
- profesia;
- calitatea produselor.

2.1. **Serii statistice**

Analiza în raport cu ***o singură caracteristică*** ne conduce la ***o serie de perechi de valori,*** care se numeşte serie statistică.

Primul număr al perechii de valori reprezintă ***valoarea caracteristicii*** iar cel de al doilea reprezintă numărul de unităţi statistice corespunzătoare acelei valori a caracteristicii.

Exemplu:

În urma unei anchete statistice în cadrul unei firme, cu privire la „productivitate“ (unităţi monetare /oră), s-au obţinut următoarele date:

Productivitatea u.m/oră	Nr. de salariaţi
50.000	2
60.000	3
70.000	4
80.000	10
90.000	6
100.000	5

$\Leftrightarrow$

x_i Val. caracter.	n_i Frecvenţa absolută
50.000	2
60.000	3
70.000	4
80.000	10
90.000	6
100.000	5

Scanned with OKEN Scanner

Observaţie :

În cazul seriilor statistice cu o singură caracteristică, pentru obţinerea unor concluzii generale, sunt eficiente 20 de valori pe care le ia caracteristica. Dacă numărul este depăşit se impune o grupare în *„clase de valori"*.

Stabilirea numărului de clase (de grupe)

- primă propunere pentru numărul de clase este dată de formula lui D. V. Huntsbergers,

$$K = 1 + 3{,}3 \cdot \lg(n), \qquad (2.1)$$

unde: K — numărul de grupe
n — nr. total de observaţii
lg(n) — logaritm zecimal din n.

În calcule se ia: partea întreagă a lui 3,3·lg(n), având :

$$K = 1 + [3{,}3 \cdot \lg(n)]$$

- a doua propunere este formla Brooks-Carruthers :

$$K < 5 \cdot \lg(n) \qquad (2.2)$$

- A treia propunere avansată de Croxton şi Cowden este ca numărul de intervale $K \in [6, 16]$.

Pentru grupare cu intervale egale, mărimea fiecărui interval va fi :

$$h = \frac{x^{sup} - x^{inf}}{K} \qquad (2.3)$$

(h – amplitudinea unei grupe).

Frecvenţe

Deosebim trei tipuri de frecvenţă :

- frecvenţa absolută notată n_i ; n_{x_i} ; $n(x_i)$;
- frecvenţa relativă f_i ; f_{x_i} ; $f(x_i)$;
- frecvenţa cumulată.

Definiţii :

Efectivul total al unei populaţii statistice este numărul tuturor elementelor acelei populaţii statistice. Se notează cu „n".

- **Frecvenţa absolută** a unei valori x_i a caracteristicii este numărul de unităţi ale populaţiei corespunzătoare acelei valori x_i. Se notează cu „n_i".

Scanned with OKEN Scanner

Suma frecvențelor absolute ale tuturor valorilor caracteristicii este egală cu efectivul total al populației.

$$\sum_{i=1}^{m} n_i = n_1 + n_2 + ... + n_m = n$$

- Se numește **frecvența relativă** (sau simplu frecvența) a unei valori a caracteristicii, raportul dintre frecvența absolută a valorii x_i (egală cu n_i) și efectivul total al populației :

$$f_i = \frac{n_i}{n} \qquad (2.4)$$

Observație :

În tabelele statistice cu două coloane, coloana a doua cu frecvențele absolute (n_i), se poate înlocui cu frecvențele relative (f_i), unde :

$$\sum_{i=1}^{m} f_i = 1 .$$

Definiție :

În loc de **caracteristică** se spune „**variabilă statistică**“ sau pe scurt variabilă.

Valoarea variabilei (caracteristicii) x_i	Frecvența absolută n_i	Frecvența relativă $f_i = \frac{n_i}{n}$
x_1	n_1	f_1
x_2	n_2	f_2
.	.	.
.	.	.
.	.	.
x_m	n_m	f_m
	$\sum_{i=1}^{m} n_i = n$	$\sum_{i=1}^{m} f_i = 1$

Frecvența definește **distribuția** sau **repartiția statistică** a variabilei X.

Scanned with OKEN Scanner

Se mai **definesc** :

- FACC — frecvența absolută cumulată crescătoare ;
- FACD — frecvența absolută cumulată descrescătoare ;
- FRCC — frecvența relativă cumulată crescătoare ;
- FRCD — frecvența relativă cumulată descrescătoare.

Definiții :

FACC a unei valori x_i a variabilei X este suma frecvențelor absolute ale tuturor valorilor variabilei care apar până la x_i inclusiv.

FACD a unei valori x_j a variabilei X este suma frecvențelor absolute ale tuturor valorilor care apar de la x_j inclusiv până la ultima valoare a variabilei.

Exemplu :

Variabila x_i	F.A n_i	FACC	FACD
x_1	n_1	n_1	$n_1 + n_2 + n_3 + n_4$
x_2	n_2	$n_1 + n_2$	$n_2 + n_3 + n_4$
x_3	n_3	$n_1 + n_2 + n_3$	$n_3 + n_4$
x_4	n_4	$n_1 + n_2 + n_3 + n_4$	n_4

Se mai spune : FRCC = FCC = FC.

(frecvența relativă cumulată crescătoare = frecvența cumulată crescătoare = frecvența cumulată).

Observații :

1. Modelul matematic al F.C. este funcția de repartiție.
2. Noțiunile introduse în raport cu valorile individuale ale variabilei pot fi extinse și în cazul tabelelor cu clase de valori.
3. **Câmpul de variație a datelor** reprezintă diferența dintre valoarea cea mai mare și valoarea cea mai mică a variabilei (caracteristicii).

Scanned with OKEN Scanner

Graficele statistice

Cele mai reprezentative grafice sunt :

- HISTOGRAMELE
- Poligoanele de frecvență
- Diagramele de structură.

a) **Histograma** este graficul obținut prin prezentarea pe abscisă, la intervale egale, a valorilor caracteristicii (valorile variabilei X) iar pe ordonată se înregistrează frecvențele (absolute sau relative).

Se construiesc dreptunghiuri (coloane) care au înălțimi proporționale cu frecvențele corespunzătoare.

b) **Poligonul de frecvențe** se construieşte pe histogramă sau fără ajutorul histogramei. În primul caz, se unesc prin segmente de dreaptă punctele centrale ale vârfurilor dreptunghiurilor histogramei.

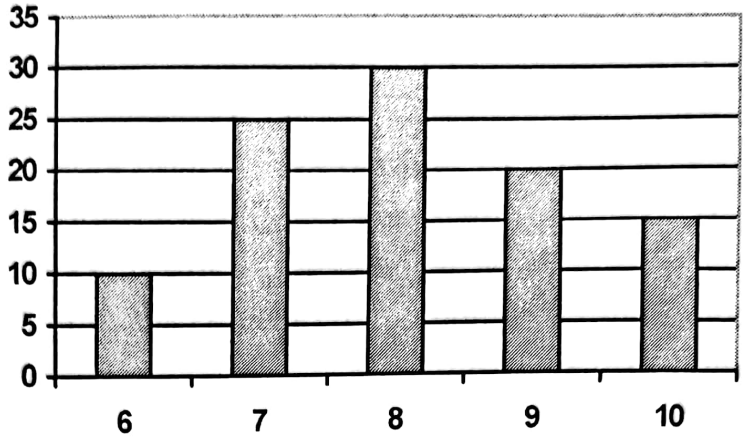
c) **Diagramele de structură** au forma unor figuri geometrice simple : cerc, pătrat, dreptunghi sau figuri compuse şi se utilizează pentru reprezentarea structurilor (frecvențelor relative (ponderi) sau procente).

Aplicație :
Se consideră seria statistică :

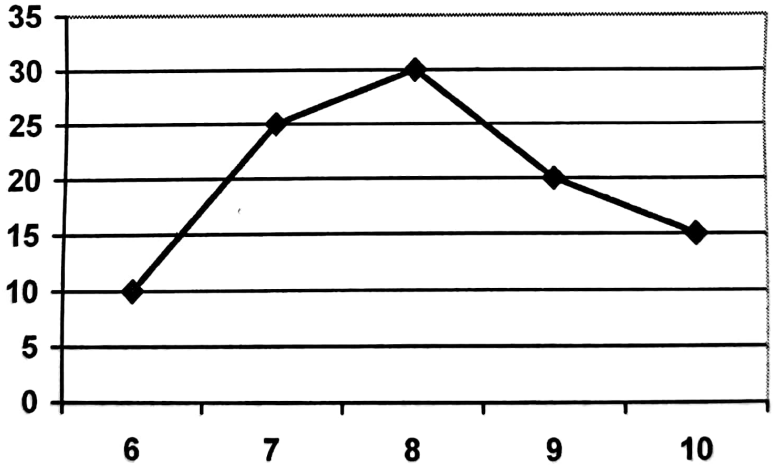
Valoarea caracteristicii x_i	Frecvența absolută n_i	F.R.	Procente %
6	10	0,10	10%
7	25	0,25	25%
8	30	0,30	30%
9	20	0,20	20%
10	15	0,15	15%
	$\sum_{i=1}^{m} n_i = 100$		

Scanned with OKEN Scanner

Histograma.



Poligonul de frecvențe *(fără ajutorul histogramei).*



Diagramă sub formă de cerc

La cerc se reprezintă proporțional frecvențele sub forma unor sectoare de cerc, ținând seama că la efectivul total corespund 360°.

Scanned with OKEN Scanner

2.2. Indicatori numerici (parametri) pentru serii statistice cu o singură caracteristică

Deosebim :

1. **Parametrii tendinței centrale (indicatori de poziție).**
2. **Parametrii împrăştierii (Indicatori ai variației observațiilor).**
3. Parametrii asimetriei şi boltirii (Indicatorii asimetriei şi boltirii).

I. Parametrii tendinței centrale (Indicatorii de poziție)

1. MEDIA
   - 1.1. Media aritmetică ($\overline{x}_a$).
   - 1.2. Media geometrică ($\overline{x}_g$).
   - 1.3. Media armonică ($\overline{x}_h$).
   - 1.4. Media pătratică ($\overline{x}_p$).
2. Mediana Me
3. MODUL (Moda dominantă) Mo
4. Valoarea centrală.
5. Momente.
6. Cuantile.

II. Parametrii împrăştierii

1. Amplitudinea.
2. Abaterea medie absolută.
3. Dispersia.
4. Abaterea medie pătratică.
5. Momentele centrale de ordin 2, 3 şi 4.
6. Coeficientul de variație.

Scanned with OKEN Scanner

III. Parametrii asimetriei şi boltirii

1. Coeficientul de asimetrie al lui PEARSON.
2. Coeficientul de asimetrie pe momentele centrale.
3. Coeficientul de boltire şi curtozis.
4. Excesul.

Serii statistice

1.Serii statistice simple.

x_i Valoarea caracteristicii	n_i Frecvența absolută
x_1	1
x_2	1
.	.
.	.
.	.
x_n	1

Efectivul total al seriei statistice n.

2. Serii statistice de frecvențe.

i	Variabila aleatoare X Valoarea caracteristicii	Frecvența absolută
	x_i	n_i
1	x_1	n_1
2	x_2	n_2
.	.	.
.	.	.
.	.	.
m	x_m	n_m

Efectivul total al seriei statistice de frecvențe :

$$n = \sum_{i=1}^{m} n_i$$

Scanned with OKEN Scanner

3. Serii statistice de frecvențe cu intervale pentru valorile caracteristicii.

i	V.a. X Valoarea caracteristicii x_i	Frecvența absoltă n_i
1	$[x_{11} \div x_{12})$	n_1
2	$[x_{21} \div x_{22})$	n_2
.	.	.
.	.	.
.	.	.
m	$[x_{m1} \div x_{m2}]$	n_m

Efectivul total al seriei statistice.

$$n = \sum_{i=1}^{m} n_i$$

Această serie statistică de frecvențe cu intervale de variație a valorii caracteristicii se înlocuieşte cu una echivalentă.

x_i	n_i
x_1	n_1
x_2	n_2
.	.
.	.
.	.
x_m	n_m

unde: x_1 este mijlocul intervalului (x_{11}, x_{12})

x_2 este mijlocul intervalului (x_{21}, x_{22})

.

.

.

x_m este mijlocul intervalului (x_{m1}, x_{m2})

Scanned with OKEN Scanner

I. *Parametrii tendinței centrale*

1. Media seriei statistice $\overline{x}$

1.1. Media aritmetică ponderată $\overline{x}_a$:

$$\overline{x}_a = \frac{\sum_{i=1}^{m} n_i x_i}{\sum_{i=1}^{m} n_i} \qquad (2.5)$$

1.2. Media geometrică ponderată $\overline{x}_G$:

$$\overline{x}_G = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_m^{n_m}} = x_1^{\frac{n_1}{n}} \cdot x_2^{\frac{n_2}{n}} \cdot \ldots \cdot x_m^{\frac{n_m}{n}} \qquad (2.6)$$

1.3. Media armonică ponderată $\overline{x}_H$:

$$\overline{x}_H = \frac{\sum_{i=1}^{m} n_i}{\sum_{i=1}^{m} \frac{n_i}{x_i}} \qquad (2.7)$$

1.4. Media pătratică ponderată $\overline{x}_p$:

$$\overline{x}_p = \sqrt{\frac{\sum_{i=1}^{m} n_i x_i^2}{\sum_{i=1}^{m} n_i}} \qquad (2.8)$$

$(\overline{x}_H < \overline{x}_G < \overline{x}_a < \overline{x}_p)$

Scanned with OKEN Scanner

2. Mediana M_e

Este valoarea caracteristicii care ocupă locul central în cadrul seriei ordonate crescător sau descrescător, împărțind seria în două părți egale.

a) **cazul seriilor simple.**

Fie seria simplă :

$x_1, x_2, ..., x_n$ (ordonată crescător) și rangurile :

1 2 n

- Dacă seria are un număr impar de termeni, mediana este acea valoare a caracteristicii cu rangul $\frac{n+1}{2}$.

Exemplu :

Cinci firme au cifrele de afaceri :

100 mil. u.m.
140 mil. u.m.
80 mil. u.m.
90 mil. u.m.
120 mil. u.m.

Se cere mediana seriei cu caracteristica: „cifra de afaceri".

Rezolvare :

Ordonăm crescător seria :

	80	90	100	120	140	(mil. u.m.)
rang (Rg)	1	2	3	4	5	

$$n = 5 \rightarrow Rg\,M_e = \frac{n+1}{2} = \frac{5+1}{2} = 3 \Rightarrow M_e = 100 \text{ mil. u.m.}$$

- Dacă seria are un număr par de termeni, mediana este dată de semisuma termenilor centrali, după ordonarea prealabilă.

Scanned with OKEN Scanner

Exemplu :

Fie seria caracteristică :

x_i	60,	80,	90,	100,	120,	130,	140,	160
Rg	1	2	3	4	5	6	7	8
x_i								

n = 8 ; termenii centrali : $Rg_1 = \frac{n}{2}$, $Rg = \frac{n}{2} + 1$ (respectiv 4 şi 5)

$$\Rightarrow M_e = \frac{100+120}{2} = 110$$

b) **Serii de frecvențe.**

i	x_i	n_i	FACC
1	2	3	3
2	5	4	7
3	10	5	12
4	12	3	15
5	15	2	17
6	20	1	18

Efectivul total:

$$n = \sum_{i=1}^{6} n_i = 3+4+5+3+2+1 = 18$$

$$\frac{n}{2} = \frac{18}{2} = 9$$

$$\frac{n}{2} + 1 = 10$$

Seria ordonată :

$x_1, x_2, \ldots x_{18}$, termenii centrali x_9, x_{10}.

Pe coloana FACC (frecvența absolută cumulată crescător) pentru x_9, x_{10} corespund la FACC = 12, valorile $x_9 = x_{10} = 10$.

Rezultă $M_e = 10$.

Altfel : 3 4 5 3 2 1

2, 2, 2; 5, 5, 5, 5; 10, 10, 10, 10, 10; 12, 12, 12; 15, 15; 20

Ranguri: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18

Scanned with OKEN Scanner

pentru $\left.\begin{array}{l} Rg=9 \Rightarrow x_9=10 \\ Rg=10 \Rightarrow x_{10}=10 \end{array}\right\} \Rightarrow M_e=\frac{10+10}{2}=10.$

c) Serii de frecvențe cu intervale pentru valorile caracteristicii.

i	Valoarea caracteristicii x_i	Frecvența absolută n_i	F.A.C.C.
1	$[x_1 \div x_2)$	n_1	n_1
2	$[x_2 \div x_3)$	n_2	$n_1 + n_2$
3	$[x_3 \div x_4)$	n_3	$n_1 + n_2 + n_3$
4	$[x_4 \div x_5)$	n_4	$n_1 + n_2 + n_3 + n_4$
5	$[x_5 \div x_6]$	n_5	$n = \sum_{i=1}^{5} n_i$

x_i	n_i
$\frac{x_1 + x_2}{2}$	n_1
$\frac{x_2 + x_3}{2}$	n_2
$\frac{x_3 + x_4}{2}$	n_3
$\frac{x_4 + x_5}{2}$	n_4
$\frac{x_5 + x_6}{2}$	n_5

Etapa 1

- **Determinarea intervalului median.**

Dacă n = par $\frac{n}{2} \in$ FACC

n = impar $\frac{n+1}{2} \in$ FACC

Exemplu: $\frac{n}{2}$ sau $\frac{n+1}{2} \in n_1 + n_2 + n_3$ și nu aparține $(n_1 + n_2) \vee (n_1 + n_2 + n_3 + n_4)$. Intervalul median este $[x_3, x_4)$

Scanned with OKEN Scanner

- **Determinarea medianei M_e**

$$n = par \Rightarrow M_e = x_{M_e} + \frac{\frac{1}{2}\left(\sum_{i=1}^{m} n_i\right) - n_{ca}}{n_{M_e}} \cdot K \qquad (2.9)$$

$$n = impar \Rightarrow M_e = x_{M_e} + \frac{\frac{1}{2}\left[\left(\sum_{i=1}^{m} n_i\right) + 1\right] - n_{ca}}{n_{M_e}} \cdot K \qquad (2.10)$$

unde : x_{M_e} — limita inferioară a intervalului median $\left(x_{M_e} = x_3\right)$;

n_{ca} — frecvențele cumulate anterior (intervalului median) ;

K — mărimea intervalului median $(x_4 - x_3)$;

n_{M_e} — frecvența intervalului median (n_3).

Aplicație.

Fie seria statistică de frecvență cu intervale pentru valorile caracteristicii :

i	x_i	n_i	FACC
1	[20 ÷ 30)	5	5
2	[30 ÷ 40)	10	15
3	[40 ÷ 50)	15	30
4	[50 ÷ 60)	30	60
5	[60 ÷ 70)	15	75
6	[70 ÷ 80]	5	80

$\Leftrightarrow$

x_i	n_i
25	5
35	10
45	15
55	30
65	15
75	5

$$n = \sum_{i=1}^{6} n_i = 80 \text{ , } n = par, \ \frac{n}{2} = 40 \in \text{ int ervalului } [50 \div 60)$$

$n = 80 = par \Rightarrow$

$$M_e = x_{M_e} + \frac{\frac{n}{2} - n_{ca}}{n_{M_e}} \cdot K$$

Scanned with OKEN Scanner

$$\left.\begin{aligned} x_{M_e} &= 50 \\ n_{ca} &= 30 \\ K &= 60-50=10 \\ n_{M_e} &= 30 \end{aligned}\right\} \quad M_e = 50+\frac{40-30}{30}\cdot 10 = 50+3{,}3333 = 53{,}3333$$

$$\bar{x}_a = \frac{\sum_{i=1}^{6} n_i x_i}{\sum_{i=1}^{6} n_i} = \frac{5\cdot 25+10\cdot 35+15\cdot 45+30\cdot 55+15\cdot 65+5\cdot 75}{80} \Rightarrow$$

$$\Rightarrow \bar{x}_a = 51{,}875$$

Observație :

Cu cât diferența $|\bar{x}_a - M_e|$ este mai mică, cu atât media aritmetică este mai reprezentativă.

3. MODUL (Moda sau dominanta) Mo

Reprezintă valoarea caracteristicii cu frecvența cea mai mare.

În cazul seriilor statistice de frecvențe şi cu intervale pentru valoarile caracteristicii, modul Mo se calculează cu formula :

$$Mo = x_{Mo} + \frac{\Delta_1}{\Delta_1+\Delta_2}\cdot K \qquad (2.11)$$

unde :

x_{Mo} — limita inferioară a intervalului modal (intervalul cu frecvența absolută maximă).

Δ_1 — diferența între frecvența absolută a intervalului modal şi frecvența absolută a intervalului premodal.

Δ_2 — diferența dintre frecvența absolută a intervalului modal şi frecvența absolută a intervalului postmodal.

K — mărimea intervalului modal (a valorilor caracteristicii).

Scanned with OKEN Scanner

Revenind la seria :

i		x_i	n_i
1	[20 ÷ 30)	25	5
2	[30 ÷ 40)	35	10
3	[40 ÷ 50)	45	15
4	[50 ÷ 60)	55	30
5	[60 ÷ 70)	65	15
6	[70 ÷ 80]	75	5

rezultă :

$$x_{Mo} = 50$$
$$K = 60 - 50 = 10$$
$$\Delta_1 = 30 - 15 = 15$$
$$\Delta_2 = 30 - 15 = 15$$
$$Mo = 50 + \frac{15}{30} \cdot 10 = 55$$

Relația între MODUL, MEDIANĂ şi MEDIA ARITMETICĂ

Metoda Pearson de determinare a modului se bazează pe relația :

$$(Mo)_p \cong 3 \cdot M_e - 2 \cdot \overline{x}_a \qquad (2.12)$$

$$(Mo)_p \cong 3 \cdot 53{,}3333 - 2 \cdot 51{,}875 = 56{,}2499$$

Valoarea reală este Mo = 55.

— eroarea absolută : Mo — $(Mo)_p = -1{,}2499$

— eroarea relativă în exprimare absolută este :

$$\frac{1{,}2499}{55} \cdot 100 = 2{,}27\%$$

Scanned with OKEN Scanner

Cuantilele

Sunt valori ale caracteristicii care împart seria statistică în „n" părți egale.

Dacă :

n = 2 cuantila este mediana ;

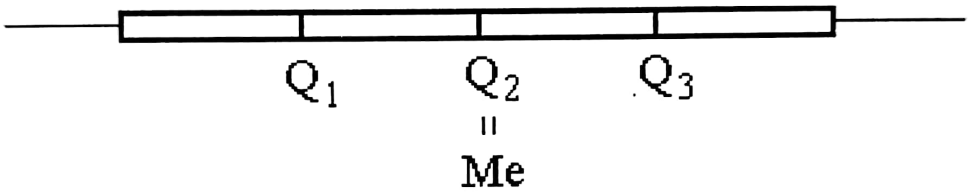
n = 4 cuantilele sunt cuartile (în număr de trei Q_1, Q_2, Q_3) ;

n = 10 cuantilele sunt decile (în număr de 9: D_1, D_2, ... D_9) ;

n = 100 cuantilele sunt centile: (în număr de 99: C_1, C_2, ..., C_{99}).

Cuartilele

Sunt în număr de trei şi împart seria în patru părți de frecvențe egale cu ¼ din efectivul total (n).



Dacă : n = par ⇒

$$Q_j = x_{Q_j} + \frac{\frac{j}{4}\left(\sum_{i=1}^{m} n_i\right) - n_{ca}}{n_{Q_j}} \cdot K, \quad j = 1,2,3 \qquad (2.13)$$

n = impar ⇒

$$Q_j = x_{Q_j} + \frac{\frac{j}{4}(n+1) - n_{ca}}{n_{Q_j}} \cdot K, \quad j = 1,2,3 \qquad (2.14)$$

unde : x_{Q_j} — este limita inferioară a intervalului în care se află

cuartila Q_k ;

n_{ca} — frecvența absolută cumulată până la intervalul în care se află cuartila Q_k ;

Scanned with OKEN Scanner

n_{Q_j} — frecvența intervalului în care se află cuartila Q_k ;

K — mărimea intervalului valorilor caracteristicii în care se află cuartila Q_k.

Revenind la exemplul anterior :

i		n_i	FACC	
1	[20 ÷ 30)	5	5	
2	[30 ÷ 40)	10	15	
3	[40 ÷ 50)	15	30	← Q_1
4	[50 ÷ 60)	30	60	← Q_2, Q_3
5	[60 ÷ 70)	15	75	
6	[70 ÷ 80]	5	80	

$\sum_{i=1}^{6} n_i = 80$ – efectivul total al seriei statistice.

Rang $Q_1 = 20 = \frac{n}{4} \in FACC = 30$.

Rang $Q_2 = 40 = \frac{n}{4} \in FACC = 60$.

Rang $Q_3 = 60 = 3\frac{n}{4} \in FACC = 60$ (sau 75).

- **Pentru cuartila Q_1 :**

$n = 80$; $x_{Q_1} = 40$; $K = 50 - 40 = 10$; $n_{ca} = 15$; $n_{Q_1} = 15$; $j = 1$;

$$Q_1 = 40 + \frac{\frac{1}{4}n - 15}{15} \cdot 10 = 40 + \frac{20-15}{15} \cdot 10 = 40 + \frac{10}{3} = 43{,}3333.$$

- **Pentru cuartila Q_2 :**

$x_{Q_2} = 50$; $K = 10$; $n_{ca} = 30$; $n_{Q_2} = 30$;

$$j = 2;\ Q_2 = 50 + \frac{\frac{2}{4} \cdot 80 - 30}{30} \cdot 10 = 50 + \frac{10}{3} = 53{,}3333$$

- **Pentru cuartila Q_3 :**

Scanned with OKEN Scanner

$x_{Q_3} = 50$; K = 10; $n_{ca} = 15$; $n_{Q_3} = 30$;

$$j = 3;\ Q_3 = 50 + \frac{\frac{3}{4} \cdot n - n_{nc}}{n_{Q_3}} \cdot K = 50 + \frac{60 - 30}{30} \cdot 10 = 60$$

Variantă :

Rg $Q_3 \in$ FACC = 75

$x_{Q_3} = 60$; K = 10 ; $n_{ca} = 60$; $n_{Q_3} = 15$;

$$Q_3 = 60 + \frac{\frac{3}{4} \cdot n - n_{nc}}{n_{Q_3}} \cdot K = 60 + \frac{60 - 60}{15} \cdot 10 = 60$$

Decilele

Sunt în număr de 9 şi împart seria în zece intervale de frecvențe egale cu 1/10 din efectivul total (n). $j = \overline{1,9}$;

$$D_j = x_{D_j} + \frac{\frac{j}{10} n - n_{ca}}{n_{D_j}} \cdot K \quad \text{ptr. n = par,} \qquad (2.15)$$

$$D_j = x_{D_j} + \frac{\frac{j}{10}(n+1) - n_{ca}}{n_{D_j}} \cdot K \quad \text{ptr. n = impar,} \qquad (2.16)$$

unde n — efectivul total al seriei ;

x_{D_j} — limita inferioară a intervalului în care se află decila Q_j,

rangul $D_j = j \cdot \frac{n}{10} \in$ FACC ;

n_{D_j} — frecvența absolută a intervalului în care se află decila D_j ;

n_{ca} — frecvența absolută cumulată anterioară intervalului în care se află decila D_j ;

K — mărimea intervalului în care se află D_j.

Scanned with OKEN Scanner

Centilele

Sunt în număr de 99 şi separă seria în 100 de intervale de frecvenţe egale.

$j = \overline{1,99}$;

$$C_j = x_{C_j} + \frac{\frac{j}{100} n - n_{ca}}{n_{C_j}} \cdot K \quad \text{pentru } n = \text{par}, \tag{2.17}$$

$$C_j = x_{C_j} + \frac{\frac{j}{100}(n+1) - n_{ca}}{n_{C_j}} \cdot K \quad \text{pentru } n = \text{impar}, \tag{2.18}$$

unde x_{C_j} — limita inferioară intervalului în care se află centila C_j,

rangul $C_j = j \cdot \frac{n}{100} \in$ FACC ;

n_{C_j} — frecvenţa absolută a intervalului ce conţine C_j ;

n_{ca} — frecvenţa absolută cumulată anterioară intervalului ce conţine C_j ;

K — mărimea intervalului ce conţine centila C_j.

II. Parametrii împrăştierii sau Indicatori ai variaţiei observaţiilor

1. Amplitudinea (A).

Reprezintă diferenţa dintre cea mai mare şi cea mai mică valoare a observaţiilor.

$$A = x_{max} - x_{min} \tag{2.19}$$

2. **Abaterea Medie Absolută (E),** sau abaterea medie liniară ($\bar{d}$). Pentru serii statistice simple (date distincte) :

Scanned with OKEN Scanner

$$E = \frac{|x_1 - \bar{x}_a| + |x_2 - \bar{x}_a| + \ldots + |x_n - \bar{x}_a|}{n} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}_a| \qquad (2.20)$$

Abaterea medie absolută (E) *reprezintă media aritmetică* ***a abaterilor de la media aritmetică*** luate în valoare absolută.

Pentru serii statistice de frecvenţe date (datele nu sunt distincte), avem :

$$\bar{d} = E = \frac{n_1|x_1 - \bar{x}_a| + n_2|x_2 - \bar{x}_a| + \ldots + n_m|x_n - \bar{x}_a|}{n_1 + n_2 + \ldots + n_m} = \frac{\sum_{i=1}^{m} n_i|x_i - \bar{x}_a|}{\sum_{i=1}^{m} n_i} =$$

$$= \frac{\sum_{i=1}^{m} f_i|x_i - \bar{x}_a|}{\sum_{i=1}^{m} f_i} = \sum_{i=1}^{m} f_i|x_i - \bar{x}_a| \qquad (2.21)$$

3. Dispersia (∇^2) (Dispersia de selecţie S^2)

Se calculează ca media aritmetică simplă sau ponderată a pătratelor abaterilor termenilor faţă de media lor.

- Pentru serii statistice simple :

$$\nabla^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_a)^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i^2}{n}\right)^2 \qquad (2.22)$$

- Pentru serii de distribuţie (de frecvenţe) :

$$\nabla^2 = \frac{\sum_{i=1}^{m} n_i(x_i - \bar{x}_a)^2}{\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m} n_i x_i^2}{\sum_{i=1}^{m} n_i} - \bar{x}_a^2, \qquad (2.23)$$

Scanned with OKEN Scanner

$$\bar{x}_a = \frac{\sum_{i=1}^{m} n_i x_i}{\sum_{i=1}^{m} n_i}$$

Observații :

1. Adeseori în locul ***dispersiei de selecție*** S^2 $\left(S^2 = \nabla^2\right)$, se folosește dispersia de selecție modificată $\tilde{S}^2$:

$$\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}_a\right)^2 = \frac{n}{n-1}S^2 \qquad (2.24)$$

2. Dispersia de selecție S^2 este o măsură a **gradului de împrăştiere** a observațiilor în jurul valorii medii a acestora ($\bar{x}_a$).

3. Modelul matematic al FRCC este funcția de repartiție.

- Populații cu distribuția normală (repartiție normală)

Distribuția normală are :

- maximul frecvenței la valoarea medie a caracteristicii;
- este simetrică în raport cu media sa;
- graficul sub formă de clopot (curba lui Gauss).

Observație :

Una din expresiile grafice ale distribuției este histograma.

4. **ABATEREA MEDIE PĂTRATICĂ (abaterea standard ∇).**

Se calculează ca medie pătratică din abaterile față de media aritmetică a valorilor caracteristicii.

- Pentru serii simple :

$$\nabla = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_a\right)^2}{n}} = \sqrt{\nabla^2} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}_a^2} \qquad (2.25)$$

- Pentru serii de distribuție (de frecvențe) :

Scanned with OKEN Scanner

$$\nabla = \sqrt{\frac{\sum_{i=1}^{m} n_i (x_i - \bar{x}_a)^2}{\sum_{i=1}^{m} n_i}} = \sqrt{\nabla^2} = \sqrt{\frac{\sum_{i=1}^{m} n_i x_i^2}{\sum_{i=1}^{m} n_i} - \bar{x}_a^2} \qquad (2.26)$$

METODA SIMPLIFICATĂ DE CALCUL A DISPERSIEI (metoda originii arbitrare)

În locul formulei

$$\nabla^2 = \frac{\sum_{i=1}^{m} n_i (x_i - \bar{x}_a)^2}{\sum_{i=1}^{m} n_i},$$

se utilizează relația

$$\nabla^2 = \frac{\sum_{i=1}^{m} n_i \left(\frac{x_i - a}{K}\right)^2}{\sum_{i=1}^{m} n_i} \cdot K^2 - (\bar{x}_a - a)^2, \qquad (2.27)$$

unde :

a — caracteristica cu frecvența cea mai mare ;

K — mărimea intervalului de grupare.

Observație.

Se mai utilizează notația :

$$x_i^* = \frac{x_i - a}{K} \text{ sau } x_i = a + Kx_i^*$$

$$\nabla^{*2} = \frac{\sum_{i=1}^{m} n_i x_i^{*2}}{\sum_{i=1}^{m} n_i} \; ; \; \bar{x}_a^* = \frac{\sum_{i=1}^{m} n_i x_i^*}{\sum_{i=1}^{m} n_i}$$

și avem $\nabla^2 = K^2 \nabla^{*2}$

5. **Coeficientul de variație (V).**

Se calculează sub două forme :

Scanned with OKEN Scanner

$$V = \frac{\nabla}{\bar{x}_a} \cdot 100 \qquad (2.28)$$

$$V_d = \frac{\bar{d}}{\bar{x}_a} \cdot 100 \qquad (2.29)$$

unde ∇ — este abaterea standard ($\nabla = \sqrt{\nabla^2}$, ∇^2 — dispersia) ;

$\bar{d}$ (E) — abaterea medie absolută (abaterea medie liniară).

Cu cât valoarea coeficientului de variație (V) este mai apropiată de zero, cu atât variația este mai slabă, ***seria este mai omogenă*** și media aritmetică este mai reprezentativă.

Dacă valoarea coeficientului de variație este mai mare de 30÷35%, se apreciază că media aritmetică nu este suficient de reprezentativă, „datele" urmând să fie separate pe grupe după alte criterii.

6. **Variabilă standardizată (sau abatere normată).**

Mărimea:

$$z_i = \frac{x_i - \bar{x}_a}{\nabla} \qquad (2.30)$$

se numește variabilă standardizată sau abatere normată. Unitatea de măsură a abaterii normate se numește unitate standard sau valoare standard.

Scanned with OKEN Scanner

2.3. **Piramida vârstelor**

În reprezentarea grafică a seriilor bidimensionale cu o caracteristică atributivă şi una cifrică se folosesc diagrame de tipul „piramida vârstelor". Construirea piramidei se efectuează prin reprezentarea în acelaşi plan a doua histograme răsturnate.

Aplicaţie

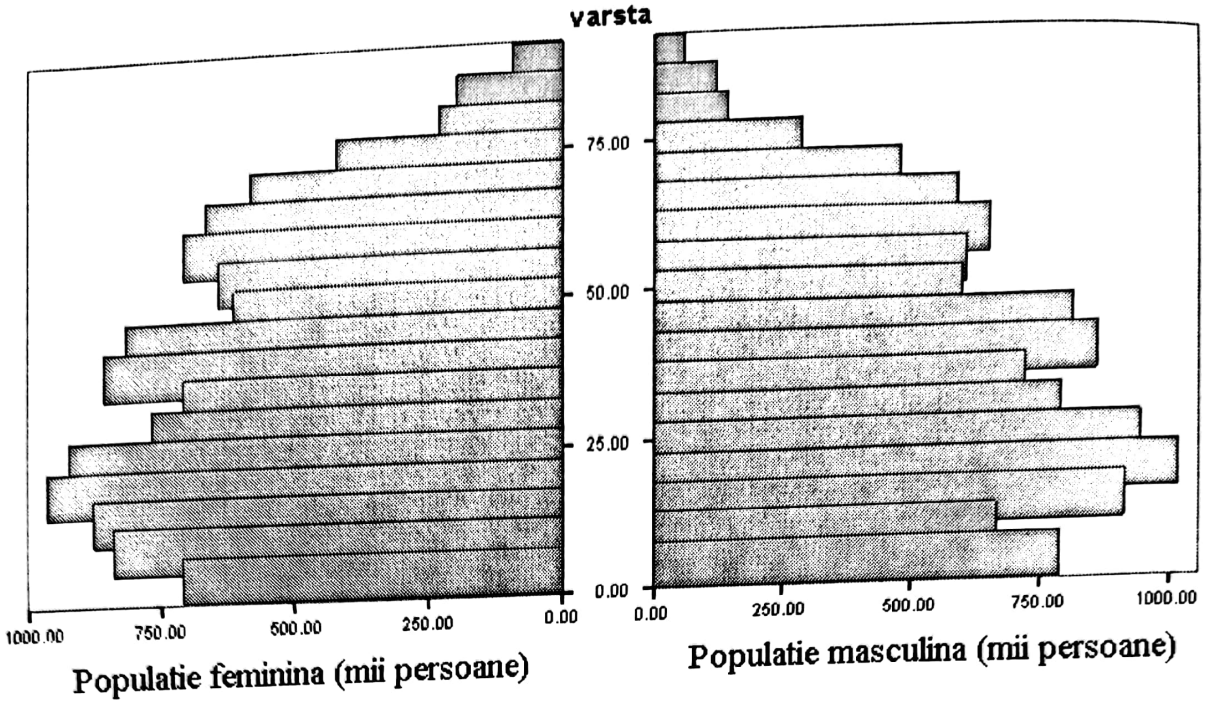
Populaţia României, după vârstă şi sex, la 1 iulie 1993 se prezenta conform datelor din următorul tabel.

Vârsta (ani)	Sex masculin (mii persoane)	Sex feminin (mii persoane)	Vârsta (ani)	Sex masculin (mii persoane)	Sex feminin (mii persoane)
0 – 4	788	708	45 – 49	598	610
5 – 9	667	838	50 – 54	603	640
10 – 14	912	875	55 – 59	646	706
15 – 19	1018	964	60 – 64	585	662
20 – 24	942	922	65 – 69	476	578
25 – 29	789	764	70 – 74	286	419
30 – 34	719	704	75 – 79	143	227
35 – 39	858	855	80 – 84	121	195
40 – 44	810	812	85 şi peste	56	94

Sursa: Anuarul statistic al României (1994)

Să se reprezinte grafic distribuţia populaţiei după vârstă şi sex.

Scanned with OKEN Scanner



Norul de puncte. Diagrama paralelipipedelor. Procedeul suprafeței poliedrale

Norul de puncte se construieşte într-un sistem cartezian şi presupune următoarele etape :

1. se trasează o rețea de linii paralele cu axele 0x şi 0y în dreptul valorilor de distribuție X şi Y ;

2. se reprezintă prin puncte frecvența x_{ij} corespunzătoare fiecărei perechi de valori (x_i, y_j).

Diagrama rezultată ia alura unui nor, denumindu-se „norul de puncte" sau diagrama de dispersie.

Procedeul paralelipipedelor şi procedeul suprafeței poliedrale reprezintă o extrapolare a histogramei şi a curbei frecvențelor în spațiul cu trei dimensiuni. Diagrama paralelipipedelor se trasează prin ridicarea unui paralelipiped cu înălțimea proporțională cu frecvența (x_{iy}) corespunzătoare perechilor de valori (x_i, y_j), pe fiecare suprafață luată ca bază. Suprafața poliedrală se construieşte prin ridicarea din centrul fiecărei suprafețe luate ca bază o linie verticală cu înălțimea proporțională cu frecvența x_{ij} şi apoi prin unirea vârfurilor acestor verticale, printr-o linie frântă sau printr-o linie curbă.

Scanned with OKEN Scanner

Aplicație.

Distribuția firmelor dintr-un eşantion, după capitalul investit x(C.I.) şi cifra de afacei y(C.A.), se prezintă în următorul tabel.

y_{j-1}, y_j C.A. mil. lei / C.I. x_{i-1}, x_i mil lei	100–300	300–500	500–700	700–900	900–1000	n_{i*}
10 – 20	1	1	–	–	–	2
20 – 30	1	1	2	3	–	7
30 – 40	–	5	10	7	–	22
40 – 50	–	1	3	5	3	12
50 – 60	–	2	2	3	2	9
60 – 70	–	1	1	2	2	6
n_{*j}	2	11	18	20	7	58

Pentru reprezentare se va aplica procedeul paralelipipedelor.

- **Diagrama polară.**

Construirea diagramei polare se realizează în următoarele etape:

1. Se reprezintă un cerc cu raza proporțională cu nivelul mediu al fenomenului.
2. se împarte cercul în perioade (de exemplu în 12 părți).
3. Se trasează raze polare pe care se reprezintă nivelele atinse de fenomen în perioadele considerate. Vârfurile razelor se unesc prin segmente de dreaptă.

Aplicație.

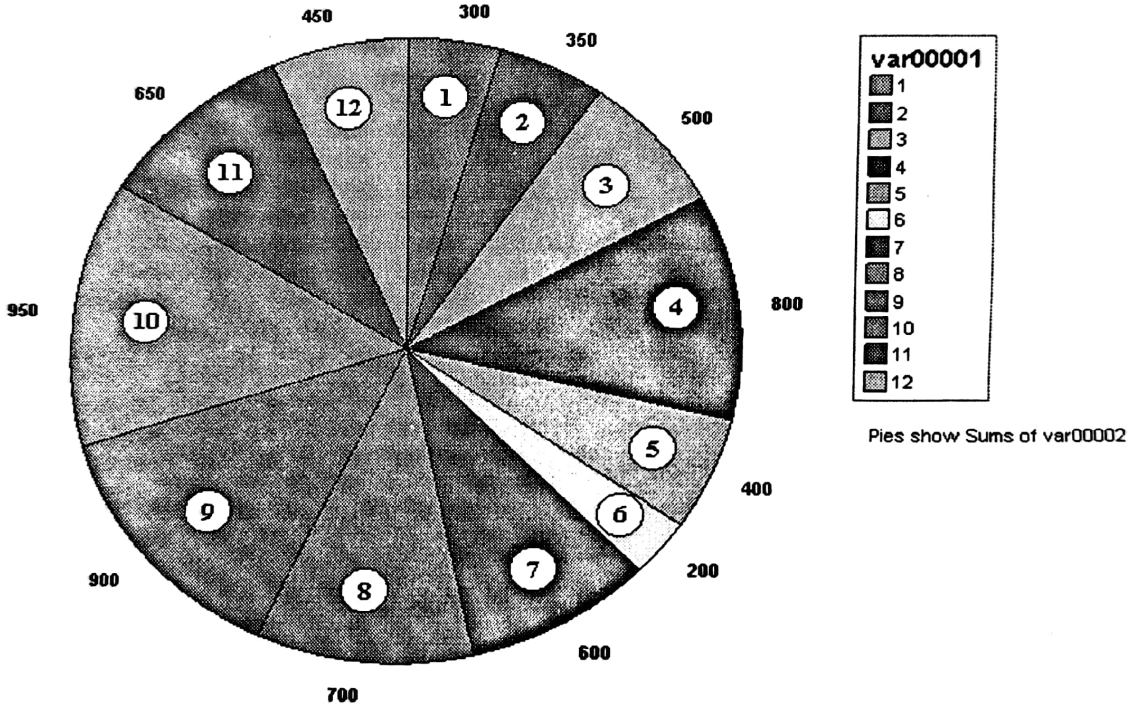
Cifra de afaceri (C.A.) a unei firme pe parcursul unui an este dată pe luni în următorul tabel.

Scanned with OKEN Scanner

Luna	1	2	3	4	5	6
C.A. mil. lei	300	350	500	800	400	200

Luna	7	8	9	10	11	12
C.A. mil. lei	600	700	900	950	650	450

Să se reprezinte în coordonate polare seria din tabel.



Scanned with OKEN Scanner

2.4. **Probleme**

Problema 2.1.

Un eşantion de firme are următoarea distribuţie după valoarea vânzărilor realizate lunar.

Grupe de firme după valoarea vânzărilor lunare realizate (milioane lei) x_i	Număr de firme n_i	Vânzări totale (milioane lei)
Sub 40	5	190
[40 – 45]	14	602
[45 – 50]	20	960
[50 – 55]	30	1568
[55 – 60]	15	870
Peste 60	6	390
TOTAL	90	4580

Să se precizeze tipul distribuţiei şi să se calculeze :

a) indicatorii tendinţei centrale ;
b) indicatorii variaţiei ;
c) indicatorii asimetriei ;
d) mărimile relative de structură şi de coordonare.

Rezolvare.

Distribuţia firmelor după valoarea vânzărilor realizate formează o serie de repartiţie unidimensională de frecvenţe cu intervale egale.

a) indicatorii tendinţei centrale.

Pentru valoarea caracteristicii alegem media fiecărui interval.

Val. C. x_i milioane lei	F.A. n_i	$x_i n_i$	$x_i \frac{n_i}{n}$	$\frac{n_i}{x_i}$	$x_i^2 n_i$
37,5	5	187,50	1,2231	0,1333	7031,50
42,5	14	595,00	1,7919	0,3294	25287,50
47,5	20	950,00	2,3583	0,4211	45125,00
52,5	30	1575,00	3,7444	0,5714	82687,50
57,5	15	862,50	1,9646	0,2609	49593,75
62,5	6	375,00	1,3174	0,0960	23437,50
TOTAL	90	4545,00	–	1,8121	233162,75

Scanned with OKEN Scanner

Mărimile medii

- media aritmetică ponderată

$$\bar{x}_a = \frac{4545}{90} = 50{,}50 \text{ milioane lei}$$

- media geometrică ponderată

$$\bar{x}_G = \sqrt[n]{x_1^{n_1}, x_2^{n_2} \ldots x_m^{n_m}} = x_1^{\frac{n_1}{n}} \cdot x_2^{\frac{n_2}{n}} \cdots x_m^{\frac{n_m}{n}} =$$

$$= 37{,}5^{\frac{5}{90}} \cdot 42{,}5^{\frac{14}{90}} \cdot 47{,}5^{\frac{20}{90}} \cdot 52{,}5^{\frac{30}{90}} \cdot 57{,}5^{\frac{15}{90}} \cdot 62{,}5^{\frac{6}{90}} = 50{,}0906 \text{ milioane lei}$$

- media armonică ponderată

$$\bar{x}_H = \frac{90}{1{,}8121} = 49{,}6661 \text{ milioane lei}$$

- media pătratică ponderată

$$\bar{x}_p = \sqrt{\frac{233162{,}75}{90}} = 50{,}8989 \text{ milioane lei}$$

Calculul indicatorilor medii de poziție

i	milioane lei	n_i	FACC	
1	sub 40	5	5	
2	[40 – 45)	14	19	← D_1
3	[45 – 50)	20	39	← Q_1
4	[50 – 55)	30	69	← Q_2, Q_3, M_0
5	[55 – 60]	15	84	← D_9
6	peste 60	6	90	

- *Cuartilele*: Q_1, Q_2, Q_3

— *Cuartila inferioară* (Q_1)

Rangul cuartilei Q_1 : $\frac{n}{4} = \frac{90}{4} = 22{,}5 = \in$ FACC = 39

Scanned with OKEN Scanner

$n = 90$; $x_{Q_1} = 45$; $k = 50 - 45 = 5$; $n_{ca} = 19$; $n_{Q_1} = 20$

$$Q_1 = x_{Q_1} + \frac{\frac{n}{4} - n_{ca}}{n_{Q_1}} \cdot k = 45 + \frac{22,5 - 19}{20} \cdot 5 = 45,8750 \text{ milioane lei}$$

– *Cuartila centrală* Q_2 : $Q_2 = M_e$ (mediana)

Rangul cuartilei Q_1 : $2 \cdot \frac{n}{4} = 45 \in FACC = 69$

$n = 90$; $x_{Q_2} = 50; k = 55 - 50 = 5; n_{ca} = 39; n_{Q_2} = 30$

$$Q_2 = x_{Q_2} + \frac{2 \cdot \frac{n}{4} - n_{ca}}{n_{Q_2}} \cdot k = 50 + \frac{45 - 39}{30} \cdot 5 = 51,00 \text{ milioane lei}$$

– *Cuartila superioară* (Q_3)

Rangul cuartilei Q_3 : $3 \cdot \frac{n}{4} = 67,50 \in FACC = 69$

$n = 90$; $x_{Q_3} = 50$; $k = 55 - 50 = 5$; $n_{ca} = 39$; $n_{Q_3} = 30$

$$Q_3 = x_{Q_3} + \frac{3 \cdot \frac{n}{4} - n_{ca}}{n_{Q_3}} \cdot k = 50 + \frac{67,50 - 39}{30} \cdot 5 = 54,75 \text{ milioane lei}$$

- **Decilele**: $D_1, D_2, D_3, \ldots, D_9$.

Împart seria în zece intervale de frecvențe egale cu $1/10$ din efectivul total al seriei. Se calculează (pentru n = par) cu formula :

$$D_j = x_{D_j} + \frac{\frac{j}{10} n - n_{ca}}{n_{D_j}} \cdot k \,, \; j = \overline{1,9}.$$

Rangul $D_j = j \cdot \frac{n}{10} \in FACC$

— **Decila D_1**

Rangul $D_1 = \frac{n}{10} = \frac{90}{10} = 9 \in FACC = 19$

Scanned with OKEN Scanner

$$n_{ca}=5, k=5; n_{D_j}=14; x_{D_1}=40$$

$$D_1=40+\frac{9-5}{14}\cdot 5=41,4286 \text{ milioane lei}$$

— **Decila D_5** : $D_5 = Q_2 = M_e = 51$ milioane lei

— **Decila D_9**

$$\text{Rangul } D_9=9\cdot\frac{n}{10}=81\in FACC=84$$

$$n_{ca} = 69 \; ; \; k = 5 \; ; \; n_{D_9} = 15 \; ; \; x_{D_9} = 55$$

$$D_9 = 55 + \frac{9\,\frac{n}{10} - 69}{15}\cdot K = 55 + \frac{81-69}{15}\cdot 5 = 59 \text{ milioane lei}$$

- **Mediala : Md**

Este indicatorul de poziție care împarte $\sum_{i=1}^{n} x_i$ sau $\sum_{i=1}^{m} x_i\, n_i$ în două părți egale. Se calculează în mod similar cu mediana.

— se determină valorile cumulate ale produselor $x_i\, n_i$

i	milioane lei	n_i	$x_i\, n_i$	$x_i\, n_i$ cumulat	
1	sub 40	5	187,50	187,50	
2	[40 – 45)	14	595,00	782,50	
3	[45 – 50)	20	950,00	1732,50	
4	[50 – 55)	30	1575,00	3307,50	← Md
5	[55 – 60]	15	862,50	4170,00	
6	peste 60	6	375,00	4545,00	

$$\text{Rangul medialei: } \frac{\sum_{i=1}^{6} x_i\, n_i}{2} = \frac{4545}{2} = 2272,50$$

$$x_{Md} = 50; \quad x_i\, n_{Md} = 1575; \quad n_{ca} = 1732,50; \; K = 5$$

Scanned with OKEN Scanner

$$Md = 50 + \frac{\frac{1}{2}\left(\sum_{i=1}^{6} x_i\, n_i\right) - n_{ca}}{x_i\, n_{Md}} \cdot K = 50 + \frac{2272{,}50 - 1732{,}50}{1575} \cdot 5$$

$$Md = 50 + 1{,}7143 = 51{,}7143 \text{ milioane lei}$$

- **Modul (dominanta) M_0**

$x_{M_0} = 50$; $k = 5$; $D_1 = 30 - 20 = 10$; $D_2 = 30 - 15 = 15$

$$M_0 = 50 + \frac{D_1}{D_1 + D_2} K = 50 + \frac{10}{25} \cdot 5 = 52 \text{ milioane lei}$$

b) **Indicatorii variației**

- Indicatorii simpli ai variației
  - Amplitudinea absolută

$$A_x = x_{max} - x_{min} = 62{,}5 - 37{,}5 = 25 \text{ milioane lei}$$

  - Amplitudinea relativă

$$A_{x\%} = \frac{A_x}{\bar{x}_a} \cdot 100 = \frac{25}{50{,}50} \cdot 100 = 49{,}5050\%$$

$$\bar{x}_a = 50{,}50$$

i	x_i milioane lei	n_i	$\lvert x_i - x_a \rvert \cdot n_i$	$(x_i - \bar{x}_a)^2 \cdot n_i$
1	37,50	5	65	845
2	42,50	14f	112	896
3	47,50	20	60	180
4	52,50	30	60	120
5	57,50	15	105	735
6	62,50	6	72	864
	TOTAL	90	474	3640

- abaterea medie absolută sau abaterea medie liniară

$$E = \bar{d} = \frac{1}{n} \sum_{i=1}^{m} \left| x_i - \bar{x}_a \right| \cdot n_i = \frac{474}{90} = 5{,}2667$$

- abaterea minimă absolută şi relativă

Scanned with OKEN Scanner

$$d_{min} = \left| x_{min} - \bar{x}_a \right| = |37,50 - 50,50| = +13 \text{ milioane lei}$$

$$d_{min\,\%} = \frac{d_{min}}{\bar{x}} \cdot 100 = -25,7426\%$$

- abaterea maximă absolută şi relativă

$$d_{max} = x_{max} - \bar{x}_a = 62,50 - 50,50 = 12 \text{ milioane lei}$$

$$d_{max\,\%} = \frac{d_{max}}{\bar{x}_a} \cdot 100 = 23,7624\%$$

Indicatori sintetici ai variației

- **Dispersia** $\nabla_x^2 \left(\delta_x^2\right)$

$$\nabla_x^2 = \frac{1}{n} \sum_{i=1}^{m} \left(x_i - \bar{x}_a\right)^2 n_i = \frac{3640}{90} = 40,4444$$

$$\text{sau } \delta_x^2 = \frac{1}{n} \sum_{i=1}^{m} x_i^2 \, n_i - \left(\bar{x}_c\right)^2 = \frac{233162,75}{90} - 50,5^2 = 40,4444$$

- **Abaterea medie pătratică** $\nabla_x \left(\delta_x\right)$

$$\nabla_x = \sqrt{\nabla_x^2} = \sqrt{40,4444} = 6,3596 \text{ milioane lei}$$

- **Coeficientul de variație (V)**

Se calculează în două forme

$$V = \frac{\nabla_x}{\bar{x}_a} \cdot 100 = \frac{6,3596}{50,50} \cdot 100 = 12,5933\%$$

$$V_d = \frac{\bar{d}}{\bar{x}_a} \cdot 100 = \frac{5,2667}{50,50} \cdot 100 = 10,4291\%$$

Se observă că V < 35% şi tinde spre zero ; variația vânzărilor este mică rezultând concluziile :

— media aritmetică ponderată ($\bar{x}_a = 50,50$ milioane lei) este o valoare reprezentativă pentru eşantion ;

— eşantionul firmelor distribuit după valoarea vânzărilor formează o colectivitate omogenă.

— gruparea firmelor după vânzări este corect efectuată ;

Scanned with OKEN Scanner

- **Indicatorii variației intercuartilice și interdecilice**

— Abaterea intercuartilică Q_c

$$Q_c = \frac{Q_3 - Q_1}{2} = \frac{54,75 - 45,8750}{2} = 4,4375 \text{ milioane lei}$$

— Coeficientul de variație intercuartilică

$$V_c = \frac{Q_c}{M_c} \cdot 100 = \frac{4,4375}{51} \cdot 100 = 8,7010\%$$

— Abaterea interdecilică Q_d

$$Q_d = \frac{D_9 - D_1}{2} \cdot 1 = \frac{59 - 41,4286}{2} = 8,7857 \text{ milioane lei}$$

— Coeficientul de variație interdecilică

$$V_d = \frac{Q_d}{M_e} \cdot 100 = \frac{8,7857}{51} \cdot 100 = 17,2269\%$$

c) **Caracterizarea și măsurarea asimetriei eșantionului**

— *Coeficientul de asimetrie a lui Pearson*

Se exprimă sub două forme:

$$C_{as} = \frac{\bar{x} - M_0}{\nabla_x}; \qquad C'_{as} = \frac{3(\bar{x} - M_e)}{\nabla_x}$$

$$C_{as} = \frac{50,50 - 52}{6,3596} = -0,2359$$

$$C'_{as} = \frac{3(50,50 - 51)}{6,3596} = -0,2359$$

mărimi ce exprimă o asimetrie negativă

— *Coeficientul intercuartilic Yule Cay*

$$C_{ay} = \frac{(M_e - Q_1) - (Q_3 - M_e)}{(M_e - Q_1) + (Q_3 - M_e)} = \frac{(51 - 45,8750) - (54,75 - 51)}{(51 - 45,8750) + (54,75 - 51)}$$

$$C_{ay} = 0,1549$$

d) **Pentru calculul mărimilor relative de structură și de coordonare se întocmește următorul tabel :**

Scanned with OKEN Scanner

i	Grupe de firme după valoarea vânzărilor (milioane lei)	Număr de firme n_i	Vânzări totale Y_i	Structura vânzărilor totale Y_i^* (% față de total)	Raportul de coordonare față de grupa VI		Vânzări medii milioane lei
					Nr. firme	Vânzări totale	
1.	sub 40 [35 – 40)	5	190	4,15	0,83	0,49	38,00
2	[40 – 45)	14	602	13,14	2,33	1,54	43,00
3	[45 – 50)	20	960	20,96	3,33	2,46	48,00
4	[50 – 55)	30	1568	34,24	5,00	4,02	52,27
5	[55 – 60]	15	870	19,00	2,50	2,23	58,00
6	peste 60 (60 – 65]	6	390	8,51	1,00	1,00	65,00
–	**TOTAL**	90	4580	100,00	–	–	–

Dacă notăm cu Y_i — vânzările totale pe grupe atunci greutatea specifică pe grupe va fi $Y_i^* = \left(Y_i \Big| \sum_{i=1}^{6} Y_i \right) \cdot 100\%$. Mărimile relative de coordonare se obţin prin compararea sub formă de raport a numărului de firme şi a vânzărilor totale.

Problema 2.2.

Distribuţia după vârsta persoanelor ce îşi desfăşoară activitatea la o întreprindere este dată în următorul tabel :

Grupe de vârstă (ani)	[32–34)	[34–36)	[36–38)	[38–40)	[40–42)	[42–44)	[44–46]
Număr de persoane	6	12	23	30	13	11	5

Scanned with OKEN Scanner

Să se calculeze :

a) media aritmetică ponderată $\bar{x}_a$
b) cuantilele Q_1, Q_2, Q_3
c) mediana M_e şi modul M_0
d) abaterea medie absolută E
e) abaterea medie pătratică ∇_x

Răspuns :

a) $\bar{x}_a$ = 38,7 ani
b) Q_1 = 36,63; Q_2 = 38,63; Q_3 = 40,65 ani
c) M_e = 38,63; M_0 = 38,58 ani
d) E = 2,354 ani
e) ∇_x = 8,99 ani.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 3

ELEMENTE DE TEORIA PROBABILITĂȚILOR. VARIABILE ALEATOARE

3.1. Teoria probabilităților

După rezultatul obținut în urma unei experiențe deosebim :

- experiențe deterministe, la care rezultatul este unic determinat, adică previzibil
- experiențe aleatoare, la care rezultatul face parte dintr-o mulțime de rezultate posibile.

Definiții :

- Orice reluare a experienței se numeşte PROBĂ (orice experiență se poate repeta de mai multe ori fără ca rezultatul să fie obligatoriu acelaşi).
- Evenimentul este rezultatul unei experiențe.
- Un eveniment elementar este un rezultat posibil al unei experiențe.

Notații :

- E = mulțimea tuturor evenimentelor elementare sau compuse asociate unei experiențe cu un număr finit de rezultate posibile.
- Ω = evenimentul sigur, ce se realizează în orice probă a experienței.

Dacă evenimentele elementare din E sunt $A_1, A_2, \ldots,$ An, atunci evenimentul sigur se poate scrie $\Omega = A_1 \cup A_2 \cup \ldots \cup An$.

- $\emptyset$ – evenimentul imposibil, care nu se realizează în nici-o probă a experienței A = eveniment; $\bar{A}$ = evenimentul opus.

Definiția clasică a probabilității

Dacă m este numărul cazurilor favorabile realizării evenimentului A, iar n este numărul cazurilor egal posibile, atunci se numeşte *probabilitate* în sens clasic a evenimentului A, numărul

$$P(A) = \frac{m}{n}$$

Proprietăți :

1. $P(\Omega) = 1$
2. $P(A \cup B) = P(A) + P(B)$, dacă A şi B sunt incompatibile adică $A \cap B = \emptyset$)
3. $P(\emptyset) = 0$
4. $P(\bar{A}) = 1 - P(A)$

Observații. Evenimentele elementare sunt echiprobabile (au aceeaşi probabilitate şi anume $\frac{1}{n}$)

Definiția axiomatică a probabilității.

Fie (Ω,K) un câmp finit de evenimente. Se numeşte probabilitate pe acest câmp o funcție de mulțimi $P:K \to R$, care satisface axiomele

1. $P(A) \geq 0$ pentru orice $A \in K$
2. $P(\Omega) = 1$
3. P este funcția finit aditivă, adică pentru $A_1, A_2 \in K$, cu $A_1 \cap A_2 = \emptyset$ are loc egalitatea : $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

Observații. Prin inducție

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

Câmp finit de probabilitate (Ω,K,P)

Proprietăți :

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) +$$
$$+ \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_K) + \ldots + (-1)^{n-1} P(\bigcap_{i=1}^{n} A_i)$$

Evenimente independente

— Evenimentele A şi B se numesc P — independente, sau independente, dacă,

$$P(A \cap B) = P(A) \bullet P(B)$$

— Probabilitate condiționată de evenimentul A, a evenimentului B.

Scanned with OKEN Scanner

Fie P(A)>0. Se numeşte probabilitate condiţionată de evenimentul A, a evenimentului B∈ K (A∈ K) notată prin :

$$P(B/A) \text{ sau } P_A(B) = \frac{P(A \cap B)}{P(A)}$$

Scheme clasice de calcul al probabilităţii

1. Schema lui Poisson

Se consideră n urne ce conţin în diferite proporţii bile albe şi negre de aceeaşi mărime şi greutate. Probabilitatea de a extrage din urna i o bilă albă este p_i iar o bilă neagră $q_i(q_i = 1 - p_i)$, $i = \overline{1,n}$. Se extrage la întâmplare câte o bilă din fiecare urnă.

Probabilitatea ca din cele n bile extrase x să fie albe este egală cu coeficientul lui t^x din dezvoltarea polinomului de gradul n

$$P_n(t) = (p_1 t + q_1)(p_2 t + q_2)\ldots(p_n t + q_n)$$

2. Schema lui Bernoulli (a bilei revenite)

Se consideră o urnă cu bile albe şi negre de aceeaşi mărime şi formă. Notăm cu p probabilitatea de a extrage o bilă albă şi cu q = 1– p probabilitatea de a extrage o bilă neagră.

Se fac n extrageri succesive, la întâmplare, cu reîntoarcerea bilei în urnă, înaintea următoarei extrageri.

Probabilitatea ca din cele n bile extrase să fie x albe este

$$P_n(x) = C_n^x \cdot p^x \cdot q^{n-x}$$

$$\text{unde } C_n^k = \frac{n!}{k!(n-k)!}; n! = 1.2.3\ldots(n-1)\cdot n$$

$$1! = 1; 0! = 1$$

Scanned with OKEN Scanner

3. Schema multinominală

Este o generalizare a schemei lui Bernoulli. În urnă se află bile de m culori, pentru care probabilitatea de a extrage o bilă de culoare i este $p_i, i = \overline{1,m}$. Se fac n extrageri succesive, la întâmplare, cu reîntoarcerea bilei după fiecare extracție.

Probabilitatea ca din cele n bile extrase x_i să fie de culoarea i (i $= \overline{1,m}$.) este $P_n(x_1,x_2,\ldots,x_m) = \frac{n!}{x_1!x_2!\ldots x_m!} p_1^{x_1} \cdot p_2^{x_2} \cdots p_m^{x_m}$

4. Schema urnei cu bilă nerevenită.

Fie o urnă cu N bile din care (a) sunt albe şi (N–a) negre.

Se extrag succesiv, la întâmplare, n bile (n<N) fără reîntoarcerea în urnă înaintea următoarei extrageri, existând posibilitatea extragerii celor n bile deodată.

Probabilitatea ca din cele n bile extrase x să fie albe

$$P_n(x) = \frac{C_a^x \cdot C_{N-a}^{n-x}}{C_N^n}$$

5. Schema urnei cu bile nerevenite în cazul mai multor culori.

O urnă cu a_i bile e culoare i, i = $\overline{1,m}$ (m≥3), bilele fiind de aceeaşi mărime. Se extrag simultan n bile, $n<a_1 + a_2 +\ldots+ a_m = N$.

Probabilitatea ca din cele n bile extrase x_i să fie de culoarea i, $i = \overline{1,m}$, $x_1 + x_2 +\ldots+ x_m = n$.

Este :

$$P_n(x_1, x_2, \ldots, x_m) = \frac{C_{a_1}^{x_1} \cdot C_{a_2}^{x_2} \cdots C_{a_m}^{x_m}}{C_N^n}$$

Scanned with OKEN Scanner

6. Schema lui Pascal (geometrică).

Fie urna din schema lui Bernoulli. Se fac mai multe extrageri, revenite, până la obținerea primei bile albe.

Probabilitatea ca la extragerea x, $(x \geq 1)$ să se obțină prima bilă albă este

$$p \cdot q^{x-1}$$

Aplicație.

La un serviciu financiar sunt verificate lucrările a trei birouri, care lucrează corect în proporție de 97%, 96% şi respectiv 95%. Se alege la întâmplare, câte o lucrare din fiecare birou. Cu ce probabilitate se vor găsi două lucrări corecte?

Rezolvare :

Se aplică schema lui Poisson cu 3 urne cu :

$$p_1 = 0{,}97; q_1 = 0{,}03; p_2 = 0{,}96; q_2 = 0{,}04; p_3 = 0{,}95; q_3 = 0{,}05$$

Probabilitatea căutată este coeficientul lui t^2 din polinomul de gradul trei:

$$P_3(t) = (p_1 t + q_1)(p_2 t + q_2)(p_3 t + q_3) \rightarrow p = 0{,}11078$$

3.2. Variabile aleatoare. Noțiuni generale

Definiția. 1

Fie o mulțime $M \neq \phi$, P(M) mulțimea părților lui M, atunci $K \subset P(M)$ se numeşte corp de mulțimi dacă :

- $(\forall) A \in K$ atunci $\overline{A} \in K$;
- $(\forall) A, B \in K$ atunci $A \cup B \in K$

Definiția. 2

Fie (Ω, K) un câmp de evenimente, unde :

Scanned with OKEN Scanner

Ω se numeşte spaţiul evenimentelor elementare ;
K se numeşte corp de evenimente.
Se numeşte probabilitate pe acest câmp o funcţie de mulţimi $P : K \to R$, care satisface axiomele :

1. $P(a) \geq 0 \qquad (\forall) A \in K$;
2. $P(\Omega) = 1$
3. P este o funcţie aditivă, adică :

$A_1, A_2, ..., A_n \in K$ cu proprietăţile :

$A_i \cap A_j = \phi \qquad (\forall) \qquad i, j \in N^*$

$$P\left(\bigcup_{j=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Definiţia. 3

Un câmp finit de evenimente (Ω, K) împreună cu o probabilitate P, definită pe acest câmp, se numeşte *câmp finit de probabilitate* şi se notează prin (Ω, K, P).

Propoziţia 1

Dacă (Ω, K, P) este un câmp de probabilitate, atunci au loc proprietăţile :

1. $P(A - B) = P(A) - P(A \cap B), (\forall) A, B \in K$
2. Dacă $A, B \in K$ şi $B \subset A$, atunci :

$P(A - B) = P(A) - P(B)$ şi $P(A) \geq P(B)$

3. $P(\overline{A}) = 1 - P(A), (\forall) A \in K$
4. $P(\phi) = 0$
5. $0 \leq P(A) \leq 1, \quad (\forall) A \in K$
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

şi

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_1) - \sum_{\substack{1 \leq i, j \leq 1 \\ i \neq j}} P(A_i \cap A_j) + \sum_{\substack{1 \leq i, j, k \leq n \\ i \neq j \neq k}} P(A_i \cap A_j \cap A_k) +$$

$$... + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Scanned with OKEN Scanner

relația lui Poincaré.

1. $P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i\neq 1}^{n} P(A_i), \quad A_i / i = \overline{1,n} \in K$

2. Fie $A_1, A_2, ..., A_n \in K$, cu proprietatea că $\bigcup_{i=1}^{n} A_i \neq \phi$, atunci :

$P\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - (n-1)$ — inegalitatea lui Boole.

Definiția 4

Fie (Ω, K) un câmp borelian de evenimente cu Ω o mulțime infinită. Se numește probabilitate complet aditivă pe acest câmp o funcție de mulțimi $P : K \to R$, care satisface axiomele :

1. $P(A) \geq 0, \qquad (\forall) A \in K$
2. $P(\Omega) = 1$
3. P este complet aditivă, adică $(\forall)$ șirul $\left\{A_n / n \in / N^*\right\}$ de evenimente din K, cu $A_i \cap A_j = \phi$ pentru $i \neq j$ este adevărată relația:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Definiția 5

Un câmp (Ω, *K*) înzestrat cu o probabilitate complet aditivă se numește câmp borelian de probabilitate și se notează (Ω, K, P).

Definiția 6

Fie (Ω, K, P) un câmp de probabilitate cu P finit aditivă. Evenimentele $A_1, A_2, ..., A_n \in K$ sunt independente dacă :

$P(A_1 \cap A_2 ... \cap A_n) = P(A_1) \cdot P(A_2) ... P(A_n)$ și sunt dependente dacă :

$$P(A_1 \cap A_2 ... \cap A_n) \neq P(A_1) \cdot P(A_2) ... P(A_n)$$

Definiția 7 (Definiția variabilei aleatoare)

Fie (Ω, K, P) un câmp borelian de probabilitate. Se numește variabilă aleatoare (unidimensională) pe acest câmp o funcție $X : \Omega \to R$ cu proprietatea că $(\forall) x \in R$ are loc relația :

Scanned with OKEN Scanner

$\{\omega / \omega \in \Omega$ si $X(\omega) \langle x\} \in K$ sau $\{\omega / x(\omega) \langle x\}$ sau $\{X \langle x\}$ unde $X(\omega)$ se numeşte valoare a variabilei aleatoare X.

Definiția 8 (Definiția funcției de repartiție)

Fie (Ω, K, P) un câmp borelian de probabilitate şi fie $X : \Omega \to R$ o variabilă aleatoare. Funcția $F : R \to R$ definită prin :

$$F(x) = P(\{\omega / X(\omega) \langle x\}) = P(X \langle x)$$

se numeşte funcție de repartiție a variabilei aleatoare X.

Proprietăți ale funcției de repartiție:

1. $F(+\infty) = \lim_{x \to \infty} F(x) = 1$; $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$

2. $0 \leq F(x) \leq 1 \quad (\forall) x \in R$

Dacă $a < b$, atunci $P(\{\omega / a \leq x(\omega) \langle b\}) = F(b) - F(a)$

1. F este o funcție nedescrescătoare, adică dacă a<b, atunci $F(a) \leq F(b)$

2. este continuă la stânga $(\forall) x \in R$, adică :

$$F(x - 0) = \lim_{\substack{x_n \to x \\ x_n \langle x}} F(x_n) = F(x)$$

Definiția 9

Fie (Ω, K, P) un câmp de probabilitate. O variabilă aleatoare $X : \Omega \to R$ care ia un număr finit de valori se numeşte variabilă aleatoare simplă.

Definiția 10

Fie (Ω, K, P) un câmp borelian de probabilitate. O variabilă aleatoare $X : \Omega \to R$ se numeşte discretă dacă $X(\Omega)$ este cel mult numărabilă. Prin repartiția variabilei aleatoare discrete X se înțelege o matrice de forma:

$$\begin{pmatrix} x_1 \dots x_n \\ P_1 \dots P_n \end{pmatrix} \quad \text{sau pe scurt} \quad \begin{pmatrix} x_i \\ P_i \end{pmatrix} \quad i = \overline{1,n}$$

Funcția de repartiție F a lui X este complet determinată de probabilitățile $P_i, i = \overline{1,n}$. Astfel $F(x) = P(X \langle x) = \sum_{i=1}^{n} P_i$. În acest caz, se spune că F este de tip discret.

Scanned with OKEN Scanner

Definiția 11

Fie (Ω, K, P) un câmp borelian de probabilitate. O variabilă aleatoare $X:\Omega \to R$ se numește *continuă*, dacă $(\exists)$ o funcție $f: R \to [0, A]$ integrabilă pe R, astfel încât :

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

unde F este funcția de repartiție a lui X.

În acest caz, f(x) se numește densitate de repartiție, iar F(x) este o funcție de repartiție continuă; expresia f(x)dx se numește probabilitate elementară.

Densitatea de repartiție are următoarele proprietăți:

1. $f(x) \geq 0$,
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $F'(x) = f(x)$
4. $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$, dacă a<b.

Funcția de repartiție a unei variabile aleatoare continue este o funcție continuă.

Caracteristici numerice ale variabilelor aleatoare

Fie (Ω, K, P) un câmp borelian de probabilitate şi $X, Y: \Omega \to R$ două variabilel aleatoare.

a) O caracteristică ce măsoară tendința centrală a repartiției variabilei aleatoare X este *media*, notată prin M(X).

Dacă *X* este o variabilă aleatoare discretă cu repartiția :

$$\begin{pmatrix} x_i \\ P_i \end{pmatrix} \quad i \in I, \quad P_i \geq 0; \ (\forall) \quad i \in I \quad \text{şi} \quad \sum_{i \in I} P_i = 1 \quad \text{atunci}$$

$$M(X) = \sum_{i=1}^{n} x_i \cdot P_i$$

Scanned with OKEN Scanner

Dacă X este o variabilă aleatoare continuă cu repartiția :

$$X = \begin{pmatrix} x \\ f(x) \end{pmatrix}, \quad x \in R, \quad f(x) \geq 0, (\forall) x \in R \text{ , şi } \int_{-\infty}^{\infty} f(x)dx = 1$$

atunci :

$$M(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

În general $M(X) = \int_{-\infty}^{+\infty} x \cdot dF(x) = \int_{-\infty}^{+\infty} x \cdot f(x)dx$

Proprietățile mediei :

1. Dacă $x(\omega) = a$, $a \in R(\forall)$ $\omega \in \Omega$, atunci $M(x) = a$.
2. $M(Kx) = K \cdot M(x)$ este funcție omogenă de gradul I.
3. $M(x + y) = M(x) + M(y)$
4. $M(x \cdot y) = M(x) \cdot M(y)$

b) Media reprezintă un caz particular al momentului inițial de ordinul r, notat prin $M_r(x)$ sau $m_r(x)$ sau m_r.

$$m_r = M(X^r) = \int_{-\infty}^{\infty} x^r dF(x)$$

Dacă *X* este o variabilă aleatoare continuă, atunci :

$$m_r = \int_{-\infty}^{\infty} x^r f(x)dx$$

Dacă *X* este o variabilă aleatoare discretă, atunci :

$$m_r = \sum x_i^r P_i$$

Pentru r=1 se obține $m_1 = M(x)$.

c) Momentul inițial absolut de ordinul r al lui X este media variabilei aleatoare $|X|^r$, adică $M(|x|^r)$.

d) O caracteristică ce măsoară împrăştierea valorilor lui X față de medie este dispersia lui X, notată prin D(X).

Dacă variabila aleatoare X este discretă, atunci dispersia este :

Scanned with OKEN Scanner

$$D(X) = \sum_i (x_i - m)^2 P_i$$

Dacă X este o variabilă aleatoare continuă, atunci dispersia este :

$$D(x) = \int_{-\infty}^{\infty} (x - m)^2 f(x)dx$$

În general, $D(x) = \int_{-\infty}^{\infty} (x - m)^2 dF(x) = M\left[(x - m)^2\right]$

Proprietăți ale dispersiei :

1. Dacă $x(\omega) = a$, $(\forall)\omega \in \Omega$, $a \in R$ atunci $D(x)=0$.

Repartiția $x = \begin{pmatrix} a \\ 1 \end{pmatrix}$, $M(x) = a < \infty$ şi $D(X) = (a - a)^2 \cdot 1 = 0$.

2. $D(aX) = a^2 D(x)$, $\quad (\forall)a \in R$

Demonstrație:

$$D(ax) = M\left[(ax - M(ax))^2\right] = M\left[a^2 (x - M(x))^2\right] = a^2 M\left(X - m^2\right) = a^2 D(x)$$

3. $D(x) = M_2(x) = M^2(x)$
4. $D(X \pm Y) = D(X) \pm D(Y)$
5. $D(x) \geq 0$

e) Abaterea medie pătratică a lui *x* este $\sigma(x) = \sqrt{D(x)}$.

f) O caracteristică mai generală decât dispersia este momentul centrat de ordinul *r*, notat $\mu_r(x)$.

$$\mu_r(x) = m_r[(x - M(x))] = M\left([x - M(x)]^r\right).$$

Variabila x-M(x) se numeşte abatere față de media variabilei aleatoare X.

Pentru r=2, se obține $\mu_2(x) = D(X)$.

Definiția funcției caracteristice

Funcția $\varphi : R \to C$, definită prin $\varphi(t) = M\left(e^{itx}\right) = \int_{-\infty}^{\infty} e^{itx} dF(x)$ se numeşte funcție caracteristică a variabilei aleatoare X.

Scanned with OKEN Scanner

$$\varphi(t)=\begin{cases}\int_{-\infty}^{\infty} e^{itx}f(x)dx & \text{daca } x \text{ este continua } x=\begin{pmatrix}x\\ f(x)\end{pmatrix}, x\in R\\ \sum e^{itx}\cdot P_j, & \text{daca } x \text{ este discreta } x=\begin{pmatrix}x_j\\ P_j\end{pmatrix}, j\in J\end{cases}$$

Repartiții clasice

O variabilă aleatoare unidimensională reală este complet determinată de repartiția sa. După cum variabilele aleatoare sunt de tip discret sau de tip continuu, repartițiile lor sunt discrete sau continue.

- **Repartiții unidimensionale discrete**

1. *Repartiția binomială* este o repartiție asociată schemei lui Bernoulli (a bilei revenite), cu parametrii *n* şi *p,* unde *n* este numărul de extrageri succesive, iar *p* este probabilitatea de a extrage o bilă albă, iar q=1–p probabilitatea de a extrage o bilă neagră.

Legea de repartiție binomială este şi densitatea de repartiție şi are forma :

$$f(x)=C_n^x p^x q^{n-x},\quad x=\begin{pmatrix}x\\ C_n^x p^x q^{n-x}\end{pmatrix};\quad M(x)=\frac{\varphi'(o)}{i}\quad M(x^2)=\frac{\varphi''(o)}{i^2}$$

Proprietăți :

1. $f(x)\geq 0, \qquad (\forall)x\in N$

$$D(x)=M(x^2)-M^2(x)$$
$$\sigma(x)=\sqrt{D(x)}$$

2. $\sum f(x)=1$

Demonstrație :

$$\sum_{x=0}^{\infty} f(x)=\sum_{x=0}^{n} C_n^x p^x q^{n-x}=C_n^o p^o q^n+C_n' p' q^{n-1}+\ldots+C_n^n p^n q^o=(p+q)^n=1$$

3. $\varphi(t)=M(e^{itx})=\sum e^{itx}C_n^x p^x q^{n-x}=\sum C_n^x (pe^{it})^x q^{n-x}=(pe^{it}+q)^n$

4. $M(x)=\dfrac{\varphi'(o)}{i}$

$$\varphi''(t)=n(pe^{it}+q)^{n-1}\cdot pie^{it}=npie^{it}(pe^{it}+q)^{n-1}$$

Scanned with OKEN Scanner

$$M(x)=\frac{\varphi'(o)}{i}=\frac{npi(p+q)^{n-1}}{i}=np$$

$$\varphi'(t)=npi\left[ie^{it}(pe^{it}eq)^{n-1}+e^{it}(n-1)(pe^{it}+q)^{n-2}\cdot pi\right]\Rightarrow$$

$$\varphi''(o)=npi[i+pi(n-1)]\Rightarrow\frac{\varphi''(o)}{i^2}=\frac{npi^2(1+p^n=p)}{i^2}=$$

$$=n^2p^2+np(1-p)=n^2p^2+npq=M(x^2)$$

$$D(x)=\left[M(x^2)\right]^2-M^2(x)=n^2p^2+npq-n^2p^2=npq$$

$$\sigma(x)=\sqrt{npq}$$

2. *Repartiția hipergeometrică*

O variabilă aleatoare X are o repartiție hipergeometrică de parametrii n, a, N, dacă repartiția sa are forma :

$$x=\begin{pmatrix} x \\ \dfrac{C_a^x C_{N-a}^{n-x}}{C_N^n} \end{pmatrix}, \qquad a\in N^*, \qquad n\in N^*,$$

$$x\in K_o=\{\max(0,a+n-N)\ldots\min(n,a)\}$$

Funcția $f(x)=\dfrac{C_a^x C_{N-a}^{n-x}}{C_N^n}$ se numeşte legea de repartiție (de probabilitate) hipergeometrică şi are proprietățile :

- $f(x)\geq 0 \qquad (\forall) \qquad x\in K_o$
- $\sum\limits_{x\in K_o} f(x)=1$

$$\sum_x \frac{C_a^x C_{N-a}^{n-x}}{C_N^n}=\frac{1}{C_N^n}\sum_x C_a^x C_{N-a}^{n-x}=\frac{C_N^n}{C_N^n}=1$$

S-a utilizat : $\sum C_a^x\cdot C_b^{n-x}=C_{a+b}^x$ care rezultă din:

$(1+t)^a(1+t)^b=(1+t)^{a+b}$

$$M(x)=\sum x\cdot f(x)=\sum x\cdot\frac{C_a^x\cdot C_{N-a}^{n-x}}{C_N^n}=\frac{1}{C_N^n}\sum xC_a^xC_{N-a}^{n-x}$$

Scanned with OKEN Scanner

$$M(x) = \sum x \cdot f(x) = \sum x \cdot \frac{C_a^x \cdot C_{N-a}^{n-x}}{C_N^n} = \frac{1}{C_N^n} \sum x C_a^x C_{N-a}^{n-x}$$

Dar :

$$C_a^x = \frac{a}{x} C_{a-1}^{x-1}$$

$$M(x) = \frac{a}{C_N^n} \sum C_{a-1}^{x-1} \cdot C_{N-a}^{n-x} = \frac{a}{C_N^n} \cdot C_{N-1}^{n-1} = \frac{a}{\frac{N}{n} C_{N-1}^{n-1}} \cdot C_{N-1}^{n-1} = a \cdot \frac{a}{N} = a \cdot p$$

unde $p = \frac{a}{N}$.

$$D(x) = M(x^2) - M^2(x)$$

$$M(x^2) = \sum x^2 \cdot \frac{C_a^x C_{N-a}^{n-x}}{C_N^n}$$

- **Câmp de probabilitate finit**

Experienţe aleatoare

Aleatoriu — are sens de întâmplător. Vine din latinescul *alea* — care înseamnă zar.

- O experienţă al cărei rezultat variază la întâmplare se numeşte *experienţă aleatoare.*
- *Probă* se numeşte rezultatul unei experienţe aleatoare.
- Situaţiile legate de o experienţă şi despre care putem spune cu certitudine că s-au produs sau nu, după efectuarea experienţei, poartă numele de *evenimente.*
- Evenimentul care poate fi realizat de o probă şi numai de una se numeşte *eveniment elementar.*

Observaţie. Fiecărei experienţe i se ataşează două evenimente cu caracter special :

— evenimentul sigur şi

— evenimentul imposibil.

- *Evenimentul sigur* este un eveniment care se realizează cu certitudine la fiecare efectuare a experienţei (se notează cu litera E).
- *Evenimentul imposibil* nu se realizează la nici o efectuare a experienţei (se notează cu ϕ).

Scanned with OKEN Scanner

2. Evenimentul imposibil constă în nerealizarea evenimentului sigur, iar evenimentul sigur constă în nerealizarea evenimentului imposibil.

- *Evenimente contrare*

  Fie o urnă cu bile albe şi negre. Notăm cu A evenimentul extragerii unei bile albe şi cu B evenimentul extragerii unei bile negre.

  Nerealizarea lui A este echivalentă cu realizarea lui B, iar nerealizarea lui B este echivalentă cu realizarea lui A. Spunem că A şi B sunt evenimente contrare.

Notaţii :

- Evenimentul contrar unui eveniment A îl notăm cu $\overline{A}$ sau CA.

  Sunt evidente relaţiile :

$$\overline{\overline{A}} = A;\ \overline{E} = \phi;\ \overline{\phi} = E.$$

- *Evenimente compatibile. Evenimente incompatibile*

Evenimentele A şi B se numesc compatibile dacă se pot realiza simultan, adică dacă există probe care realizează atât pe A, cât şi B. În caz contrar, evenimentele sunt incompatibile.

Observaţie. Evenimentele contrare sunt incompatibile.

- *Eveniment implicat de alt eveniment*

Definiţie. Evenimentul B este implicat de evenimentul A, sau evenimentul A implică evenimentul B, dacă B se realizează de fiecare dată când se realizează A.

- Orice eveniment implică evenimentul sigur.

Densitate de repartiţie

Definiţie. Fie $X : \Omega \to R$ o variabilă aleatoare şi $F(x)$ funcţia ei de repartiţie.

- Dacă există o funcţie $f : R \to R$ are proprietăţile :

  a) $f(x) \geq 0 \qquad (\forall)x \in R$

  b) $F(x) = \int_{-a}^{x} f(y)dy, \qquad (\forall)x \in R$

atunci se spune :

1. funcţia f(x) este *densitate de repartiţie* (sau *de probabilitate*)
2. variabila aleatoare X este absolut continuă.

Scanned with OKEN Scanner

Observație.

Densitatea de repartiție trebuie să verifice condițiile :

a. $f(x) \geq 0$ pentru orice $x \in R$

b. $\int_{-\infty}^{+\infty} f(x)dx = 1$

- $P(x_1 \leq x \langle x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(y)dy$

3.3. **Distribuții (Repartiții) discrete și continue. Proprietăți.**

$$\text{Fie :} \quad X:\begin{pmatrix} x_1 x_2 \ldots x_m \\ p_1 p_2 \ldots p_m \end{pmatrix}; \qquad Y:\begin{pmatrix} y_1 y_2 \ldots y_n \\ q_1 q_2 \ldots q_n \end{pmatrix} \qquad (3.1)$$

- Pe prima linie se trec *valorile posibile* ale *v.a.X* și sub fiecare valoare probabilitatea cu care *X* ia această valoare.
- Tabloul (1) definește *distribuția (repartiția) v.X.*

$$P(X = x_j) = p_j \; ; \qquad \sum_{j=1}^{m} p_j = 1 \qquad (3.2)$$

- Fiind date v.a.X, Y și o constantă a, avem:

$$aX:\begin{pmatrix} ax_1 & ax_2 \ldots ax_m \\ p_1 & p_2 \cdots p_m \end{pmatrix}$$

$$a+X:\begin{pmatrix} a+x_1 & a+x_2 & \ldots a+x_m \\ p_1 & p_2 & \cdots p_n \end{pmatrix}$$

$$X+Y:\begin{pmatrix} x_1+y_1 & x_1+y_2 & \ldots x_m+y_n \\ p_{11} & p_{12} & \cdots p_{mn} \end{pmatrix};$$

$$XY:\begin{pmatrix} x_1y_1 & x_1y_2 & \ldots x_m y_n \\ p_{11} & p_{12} & \cdots p_{mn} \end{pmatrix}$$

unde p_{ij} este *probabilitatea realizării simultane* a egalităților $X = x_i$, $Y = y_j$.

- Dacă *evenimentele* $X = x_i$, $Y = y_j$ *sunt independente* pentru toate valorile indicilor *i* și *j* ($i = \overline{1,m}; j = \overline{1,n}$), spunem că variabilele *X* și *Y* sunt independente și $p_{ij} = p_i \cdot q_j$.

Scanned with OKEN Scanner

Definiţii

Definiţia 1. O variabilă aleatoare X este continuă *dacă valorile posibile ale variabilei* reprezintă un interval al axei reale (a, b).

- *Legea de distribuţie a unei variabile* aleatoare continue se dă cu ajutorul unei funcţii $f : /R \to /R$, $f(x) \geq 0$, numită *densitate de repartiţie.*
- Probabilitatea $p(a\langle x\langle b)$ ca valoarea x a variabilei aleatoare X să aparţină intervalului (a,b) se calculează cu relaţia:

$$p(a\langle x\langle b) = \int_a^b f(x)dx \qquad (3.3)$$

cu menţiunea că $\int_{-\infty}^{+\infty} f(x)dx = 1$ (3.4)

- Dacă notăm cu F(x) probabilitatea P(X<x), atunci:

$$F(x) = \int_{-\infty}^{x} f(x)dx \qquad (3.5)$$

Definiţia 2. Valoarea medie matematică de aşteptare a unei *v.a.c.X* se calculează cu relaţia:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx \qquad (3.6)$$

iar pentru *v.a.discrete (1)* cu formula :

$$M(X) = \sum_{j=1}^{n} x_j p_j$$

Definiţia 3. Numim *moment de ordinul K* al variabilei X, valoarea medie a variabilei X^k:

$$M_K(X) = \int_{-\infty}^{+\infty} x^k f(x)dx; \qquad M_K(X) = \sum_{j=1}^{n} x_j^k p_j \qquad (3.7)$$

$$M_K(X) = M(X^k)$$

Definiţia 4 . fiind dată o v.a.c.X şi o constantă α a variabilei X, variabila X-α:

- Variabila X-M(X) este denumită *abaterea variabilei X.*
- *Împrăştierea variabilei X* se caracterizează *prin valoarea medie a abaterii absolute* $|X - M(x)|$ numită *abatere medie.*

Definiţia 5. Momentul centrat de ordinul K :

$$M_k(X) = \int_{-\infty}^{+\infty} (x - m)^k f(x)dx \qquad (3.8)$$

Scanned with OKEN Scanner

unde m=M(X), $k \in Z_+$.

Definiția 6. Dispersia unei variabile aleatoare X este momentul centrat de ordinul al doilea al acestei variabile.

$$\sigma^2 = D^2(X) = M\left[(x-m)^2\right] = M(X^2) - (M(X))^2 \qquad (3.9)$$

Definiția 7. Abaterea medie pătratică D(X) a unei variabile aleatoare:

$$\sigma = D(X) = \sqrt{D^2(X)} = \sqrt{M(X^2) - [M(X)]^2} \qquad (3.10)$$

Legi de distribuție

1. Legea de distribuție normală :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in R, \ \sigma > 0. \qquad (3.11)$$

Se notează cu F(x; m, σ) – *densitatea de repartiție normală* cu parametrii *m* şi σ, iar cu F(x,0,1) *funcția de repartiție* corespunzătoare densității de repartiție f(x; 0,1) :

$$F(x;0,1) = \int_{-\infty}^{x} f(t;0,1)dt = \frac{1}{2} + \phi(x) \qquad (3.12)$$

unde $\phi(x) = \frac{1}{2\pi}\int_0^x e^{-\frac{t^2}{2}} dt$, denumită funcția lui Laplace.

- Dacă variabila aleatoare *X* urmează legea normală cu parametrii *m* şi σ, atunci v.a. $Y = \frac{X-m}{\sigma}$ urmează legea normală cu parametrii 0 şi 1.

$$F(X) = P(Y\langle x) = P(X\langle m + \sigma x) = F(m + \sigma x; m, \sigma).$$

- Probabilitatea realizării inegalității a<X<b, dacă *X* urmează legea normală cu parametrii *m* şi σ şi $a, b \in R$, este dată de relația :

$$P(a\langle X\langle b) = F\left(\frac{b-m}{\sigma};0,1\right) - F\left(\frac{a-m}{\sigma};0,1\right) = \phi\left(\frac{b-m}{\sigma}\right) - \phi\left(\frac{a-m}{\varpi}\right) \qquad (3.13)$$

În particular, $P\left(|x-m|\langle K\sigma\right) = 2\cdot\phi(K)$ (3.14)

unde $\phi(K) = F(K) - 0,5$, F(K) fiind valoarea funcției de repartiție normală *F(x)*, a cărei repartiție este dată în (3.12).

Scanned with OKEN Scanner

2. Repartiția Student cu n grade de libertate, caracterizată prin densitatea de repartiție:

$$f_n(x) = \frac{1}{\sqrt{\pi n}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in /R$$

cu $M(X) = 0; \quad D^2(X) = \frac{n}{n-2}, \quad n>2.$

Avem : $\lim_{n\to\infty} f_n(x) = f(x;0,1)$, adică densitatea de repartiție normală cu parametrii 0 şi 1.

- În *cazul variabilelor independente* X_1, X_2, …, X_{n+1}, avem o variabilă aleatoare:

$$\frac{\sqrt{n}X_{n+1}}{\sqrt{X_1^2+X_2^2+...+X_n^2}}$$ care are distribuție Student cu n grade de libertate.

3. Repartiția Fisher cu (m, n) grade de libertate. Densitatea de repartiție este:

$$f(t) = \begin{cases} K_{mn}\, t^{\frac{n-2}{2}} (mt+n)^{-\frac{m+n}{2}}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

unde $K_{mn} = \frac{m^{\frac{m}{2}} \cdot n^{\frac{n}{2}} \cdot \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)}$ cu $M(X) = \frac{n}{n-2}, \quad n>2$;

$$D^2(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

4. *Repartiția exponențială* :

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \lambda \rangle 0 \\ 0, & x < 0 \end{cases}$$

Scanned with OKEN Scanner

$$M(X)=\frac{1}{\lambda};\qquad D^2(X)=\frac{1}{\lambda^2},$$

5. *Repartiția gama* :

$$f(x)=\begin{cases}\dfrac{\beta^{\lambda}}{\Gamma(\lambda)}x^{\lambda-1}e^{-\beta x}, & x\rangle 0,\ \lambda\rangle 0,\beta\rangle 0\\ 0, & x\langle 0\end{cases}$$

$$M(X)=\frac{1}{\beta};\qquad D^2(X)=\frac{\lambda}{\beta^2}$$

6. *Repartiția beta* cu (*m, n*) grade de libertate :

$$f(x)=\begin{cases}\dfrac{1}{B(m,n)}x^{m-1}(1-x)^{n-1}, & x\in(0,1)\\ 0, & x\notin(0,1)\end{cases}$$

$$M(X)=\frac{m}{m+n};\qquad D^2(X)=\frac{mn}{(m+n)^2(m+n+1)}$$

7. *Repartiția hi pătrat* x^2, cu *n* grade de libertate :

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\cdot\Gamma\left(\frac{n}{2}\right)}\cdot x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}, & x\geq 0\\ 0, & x\langle 0\end{cases}$$

$$M(X)=0;\qquad D^2(X)=2n$$

Scanned with OKEN Scanner

8. *Repartiția Weibull*:

$$f(x)=\begin{cases}\alpha\lambda x^{\alpha-1}\cdot e^{-cx^{\lambda}} & x\rangle 0\\ 0 & x\langle 0\end{cases},$$

$$M(X)=c^{-\frac{1}{\alpha}}\cdot\Gamma\left(1+\frac{1}{\alpha}\right)$$

$$D^2(X)=c^{\frac{2}{\alpha}}\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right]$$

Observație. Particularizând relațiile :

$$p(X\langle a)=F\left(\frac{a-m}{\sigma}\right)$$

$$p(X\rangle a)=1-F\left(\frac{a-m}{\sigma}\right) \qquad (3.16)$$

$$p(X\notin(a,b))=p(X\langle a)+p(X\rangle b)$$

unde m reprezintă valoarea medie matematică de așteptare și σ - abaterea medie pătratică a *v.a.X*.

- Dacă mărimea aleatoare X reprezintă o combinație liniară a *m.a.*X_j pentru care sunt cunoscute *v.m.m.a.*, mx_j și *abaterile medii pătratice* σx_j:

$$X=\sum_{j=1}^{n}c_jX_j,\quad \text{atunci}\quad m_x=\sum_{j=1}^{n}c_j\cdot mx_j;\sigma_x^2=\sum_{j=1}^{n}c_j^2\sigma^2x_j \qquad (3.17)$$

- Pentru dependențe neliniare ale mărimii aleatoare X, $X = f(X_1, X_2, \ldots X_n)$ în ipoteza unor abateri mici ale mărimilor X_j față de v.m.m.a. mx_j, avem:

$$m_x=f(m_{x_1},\ldots,m_{x_n});\quad \sigma_x^2=\sum_{j=1}^{n}\left(\frac{\partial f}{\partial X_j}\right)^2\sigma_{x_j}^2 \qquad (3.18)$$

- Dacă mărimile aleatoare X_j, $j=\overline{1,n}$ sunt independente și au v.m.m.a. $M(X_j)$ și dispersiile $D^2(X_j)$ limitate de același nr.C, atunci pentru $(\forall)\,\varepsilon>0$ este valabilă T.lui Cebâșev :

Scanned with OKEN Scanner

$$P\left(\left|X / M(X)\right| \langle \varepsilon\right) \geq 1 - \frac{C}{n\zeta^2} \qquad (3.19)$$

unde $X = \frac{1}{n}\sum_{j=1}^{n} X_j$; $M(X) = \frac{1}{n}\sum_{j=1}^{n} M(X_j)$ din care rezultă:

$$\lim_{n\to\infty} P\left(\left|X - M(X)\right| \rangle \varepsilon\right) = 1 \qquad (3.20)$$

Repartiții clasice

1. Repartiții discrete unidimensionale

1a. Repartiția uniformă $X : \begin{pmatrix} x_k \\ p_k \end{pmatrix}$, $p_k = \frac{1}{n}$

1b. Repartiția binomială $X : \begin{pmatrix} k \\ p_k \end{pmatrix}$, $p_k = C_n^k p^k q^{n-k}$ $p\rangle 0, q\rangle 0$ $p+q=1$

1c. Repartiția hipergeometrică

$$X : \begin{pmatrix} k \\ p_k \end{pmatrix}, \quad p_k = \frac{C_{N_1}^k \cdot C_{N_2}^{n-k}}{C_{N_1+N_2}^n} \quad n \leq N_1 + N_2 \quad k \leq N_2 \quad n - k \leq N_2$$

1d. Repartiția Poisson

$$X : \begin{pmatrix} k \\ p_k \end{pmatrix}, \qquad p_k = \frac{a^k}{k!} e^{-a}, \quad k \in / N, \qquad a\rangle 0$$

1e. Repartiția geometrică $p_k = pq^k$ $p\rangle 0, q\rangle 0$

2. Repartiții continue unidimensionale

2a. Repartiția uniformă

$$X : \begin{pmatrix} x \\ f(x) \end{pmatrix}, f(x) = \begin{cases} \frac{1}{b-a} & , x \in (a,b), \quad a\langle b \\ 0 \quad \text{in} \quad \text{rest} \end{cases}$$

2b. Repartiția exponențială (negativă)

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 \quad \text{in} \quad \text{rest} \end{cases}$$

Scanned with OKEN Scanner

2c. Repartiția gama

$$f(x)=\frac{x^{p-1}e^{-\frac{x}{q}}}{q^p\Gamma(p)},\qquad x\geq 0\quad p\rangle 0,\quad q\rangle 0$$

2d. Repartiția beta $\beta(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}dx,\quad p\rangle 0,q\rangle 0$

$$f(x)=\frac{1}{\beta(p,q)}x^{p-1}(1-x)^{q-1},\quad x\in\leq[0,1]\qquad p\rangle 0,q\rangle 0$$

2e. Repartiția Weibull

$$f(x)=\frac{\beta}{\gamma}(x-\alpha)^{\beta-1}\exp\left\{-\frac{(x-\alpha)^{\beta}}{\gamma}\right\},\quad x\geq\alpha,\alpha\geq 0,\beta\rangle 0,\gamma\rangle 0$$

2f. Repartiția normală

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-m)^2}{1\sigma^2}\right\},\quad x\in R,\ m\in R,\ \sigma\rangle 0$$

2g. Repartiția lognormală

$$f(x)=\frac{1}{\sigma x\sqrt{2\pi}}\exp\left(-\frac{1}{2\sigma^2}l_n^2\frac{x}{m}\right)\qquad x\rangle 0,\quad m\rangle 0,\quad \sigma\rangle 0$$

2h. Repartiția x^2 (Helmert-Pearson)

$$f(x)=\frac{1}{2^{\frac{n}{2}}\sigma^2\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2\sigma^2}},\ x\geq 0,\ \sigma\rangle 0,\ n\in N^*,n=2,4,6,10,20$$

2i. Repartiția Student

$$f(x)=\frac{1}{\sqrt{n}\beta\left(\frac{1}{2},\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},x\in R,\qquad n\in N^*$$

$$\beta(p,q)\cdot\Gamma(p+q)=\Gamma(p)\cdot\Gamma(q)$$

2j. Repartiția Snedecor

$$f(x,n_2,n_2)=\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}\cdot\frac{1}{\beta\left(\frac{n_1}{2},\frac{n_2}{2}\right)}\cdot x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}},\qquad x\geq 0$$

Scanned with OKEN Scanner

$n_1, n_2 \in N \qquad n_1 \rangle 1, n_2 \rangle 1$

2k. Repartiția Fischer

$$f(x) = 2\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \cdot \frac{1}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} e^{n_1 x}\left(1 + \frac{n_1}{n_2} e^{2x}\right)^{-\frac{n_1+n_2}{2}}$$

Formule şi scheme clasice de probabilitate

1. $P(A \cup B) = P(a) + P(B) - P(A \cap B)$
2. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) -$
   $- P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Demonstrația.1

Fie n numărul cazurilor egal posibile ale experienței în raport cu care A şi B sunt evenimente

m = numărul cazurilor favorabile lui A şi

s = numărul cazurilor favorabile lui B

Presupunem că din cele m cazuri favorabile lui A, t sunt favorabile lui $A \cap B$.

Numărul cazurilor favorabile lui $A \cup B$ este $m + s - t$ (şi nu $m + s$, deoarece în acest caz t cazuri ar fi numărate de 2 ori şi la A şi la B).

Rezultă :

$$P(A) = \frac{m}{n}, \qquad P(B) = \frac{s}{n}; \qquad P(A \cap B) = \frac{t}{n}$$

$$P(A \cup B) = \frac{m+s-t}{n} = \frac{m}{n} + \frac{s}{n} - \frac{t}{n} = P(A) + P(B) - P(A \cap B)$$

$$P(A_1 \cup A_2 \cup A_n \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) -$$
$$- P(A_1 \cap A_2) - \ldots - P(A_{n-1} \cap A_n) + P(A_1 \cap A_2 \cap A_n) +$$
$$+ \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Demonstrația. 2 : Prin inducție matematică.

Aplicația. 1 O urnă conține 3 bile albe şi 7 negre, iar alta 7 bile albe şi 3 bile negre. Din fiecare urnă se extrage câte o bilă. Care este probabilitatea să obținem cel puțin o bilă albă?

Scanned with OKEN Scanner

Rezolvare: Fie evenimentul extragerii unei bile alte din prima urnă şi *B* evenimentul extragerii unei bile albe din a două urnă.

$$P(A) = \frac{3}{10}, \qquad P(B) = \frac{7}{10}$$

Deoarece *A* şi *B* sunt evenimente independente

$$P(A \cap B) = P(A) \cdot P(B).$$

avem:

$$P(A \cup B(= P(A) + P(B) - P(A \cap B) = \frac{3}{10} + \frac{7}{10} - \frac{3}{10} \cdot \frac{7}{10} = 1 - \frac{21}{100} = \frac{79}{100}$$

Aplicația. 2 Problema concordanțelor

O urnă conține *n* bile numerotate 1,2,...n. Se fac extrageri succesive fără a pune bila înapoi în urnă. Care este probabilitatea obținerii a cel puțin unei concordanțe.

Observație. Spunem că am obținut o concordanță la extragerea *k*, dacă la acea extragere am obținut bila *k.*

Rezolvare: Notăm A_i (i=1,...,n) ev.obținerii unei concordanțe la extragerea de rangul *i.*

$$P(A_1(= P(A_2) = ... = P(A_n) = \frac{(n-1)}{n!} = \frac{1}{n}$$

Calculăm probabilitatea unei intersecții de două evenimente $A_i \cap A_j$. Păstrăm nemişcați doi indici, obținem (*n*-2)! grupe.

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}. \qquad \textit{În total avem } C_n^2 \textit{ intersecții de 2 ev.}$$

În general : $P(A_{i1} \cap A_{i2} \cap ... \cap A_{ik}) = \frac{1}{n(n-1)(...)(n-k+1)}$ şi avem C_n^k intersecții de *k* evenimente.

Notăm A = evenimentul obținerii unei concordanțe.

$$A = A_1 \cup A_2 \cup ... \cup A_n$$

$$P(A) = P(A_1 \cup A_2 \cup ... \cup A_n) = n \cdot \frac{1}{n} - C_n^2 \frac{1}{n(n-1)} +$$

$$+ C_n^3 \frac{1}{n(n-1)(n-2)} - ... + C_n^n \frac{(-1)^{n+1}}{n(n-1)...s}$$

Scanned with OKEN Scanner

$$P(A) = 1 - \frac{1}{1\cdot 2} + \frac{1}{1\cdot 2\cdot 3} - \frac{1}{1\cdot 2\cdot 3\cdot 4} + \ldots + (-1)^{n+1}\frac{1}{1\cdot 2\cdot \ldots \cdot n}$$

Schema lui Poisson

Se dau n urne $U_1, U_2, \ldots, U_n$ care conţin bile albe şi negre în proporţii date.

Cunoaştem probabilităţile $p_i (i = \overline{1,n})$ cu care este extrasă o bilă albă din urna U_i.

Se cere probabilitatea de a extrage k bile albe şi n-k bile negre, atunci când din fiecare urnă se extrage câte o bilă.

Rezolvare. Pentru a extrage k bile albe şi $(n-k)$ bile negre, trebuie să se realizeze k evenimente A_i şi (n-k) evenimente $\overline{A_i}$ (A_i= ev.extr.unei bile albe din urna U_i).

Evenimentul $A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik} \cap \overline{A}_{i_{k+1}} \cap \ldots \cap \overline{A}_{i_n}$ se realizează cu probabilitatea:

$$p_{i_1} \cdot p_{i_2} \ldots p_{i_k} \cdot q_{i_{k+1}} \cdot q_{i_{k+2}} \cdot \ldots \cdot q_{i_n}$$

(indicii $i_1 i_2, \ldots, i_k, i_{k+1}, \ldots i_n$ = reprezintă o permutare a indicilor 1,2,…n.

- Evenimentul căruia dorim să-i calculăm probabilitatea este reuniunea tuturor evenimentelor de forma de mai sus şi deci probabilitatea sa este egală cu suma tuturor produselor de forma:

$$p_{i_1} \cdot p_{i_2} \ldots p_{i_k} \cdot q_{i_{k+1}} \cdot q_{i_{k+2}} \cdot \ldots \cdot q_{i_n}$$

unde p apare de k ori (cu diferiţi indici), iar q de $(n-k)$ ori cu indici care nu apar la p.

Observaţie. Se observă că după aceeaşi regulă se calculează coeficientul lui x^k în polinomul :

$$P(x) = (p_1 x + q_1)(p_2 x + q_2)\ldots(p_n x + q_n)$$

Aplicaţie. Într-un atelier sunt trei maşini. Prima dă 0,9% rebuturi, a doua 1% şi a treia 1,3%. Se ia la întâmplare câte o piesă de la fiecare maşină şi se cere probabilitatea ca 2 dintre piesele luate să fie bune şi una să fie rebut.

Rezolvare. Suntem în cazul schemei lui Poisson. Probabilitatea căutată va fi coeficientul lui x^2 din produsul :

Scanned with OKEN Scanner

$$(p_1x+q_1)(p_2x+q_2)(p_3x+q_3)$$

unde $p_1=0{,}991,\quad p_2=0{,}990,\quad p_3=0{,}987$

$q_1=0{,}009,\quad q_2=0{,}010,\quad q_3=0{,}013$

Schema lui Bernoulli

Să presupunem că în schema lui Poisson urnele $U_1, U_2, ..., U_n$ sunt identice. Atunci

$$p_1=p_2=...=p_n=p;\quad q_1=q_2=...=q=1-p\ .$$

Probabilitatea extragerii a *k* bile albe şi (*n-k*) bile negre, în acest caz, va fi coeficientul lui x^k din polinomul:

$$P(x)=(px+q)^n$$

adică va fi egală cu

$$C_n^k p^k q^{n-k}$$

Observaţie. Recunoaştem în acest caz termenul general al ridicării la puterea a *n*-a a binomului $(p+q)$. Pentru acest motiv, schema se mai numeşte *binomială*.

Valori medii

Fie variabila aleatoare $X\begin{pmatrix} x_1x_2...x_n \\ p_1p_2...p_n \end{pmatrix}$.

Notăm M(X) = valoarea medie a variabilei aleatoare X.

Definiţie. $M(X)=p_1x_1=p_2x_2+...+p_nx_n=\sum_{i=1}^{n}p_ix_i$

- *Proprietăţi ale valorii medii*
- Valoarea medie a unei constante este egală cu constanta.
- Distribuţia unei variabile aleatoare care ia o singură valoare este de forma :

$\begin{pmatrix} a \\ 1 \end{pmatrix}$ şi deci valoarea medie $a\cdot 1=a$.

- Fie *X* şi o constantă *a* :
- $M(a+X)=a+M(X)$
- $M(aX)=a\cdot M(X)$

Scanned with OKEN Scanner

Demonstrație.

$$\text{Fie } X\begin{pmatrix} x_1 x_2 \ldots x_n \\ p_1 p_2 \cdots p_n \end{pmatrix};\quad a+X\begin{pmatrix} a+x_1, a+x_2 \ldots a+x_n \\ p_1 \qquad\quad p_2 \cdots p_n \end{pmatrix}$$

$$M(a+X) = p_1(a+x_1) + p_2(a+x_2) + \ldots + p_n(a+x_n) =$$

$$= a(p_1 + \ldots p_n) + p_1x_1 + p_2x_2 + \ldots + p_nx_n = a + M(X)$$

$$aX\begin{pmatrix} ax_1 \qquad ax_2 \ldots ax_n \\ p_1 \qquad p_2 \cdots \quad p_n \end{pmatrix}$$

$$M(ax) = p_1ax_1 + p_2ax_2 + \ldots + p_nax_n = a(p_1x_1 + \ldots + p_nx_n) = aM(X)$$

- Fie $X\begin{pmatrix} x_1 x_2 \ldots x_n \\ p_1 p_2 \cdots p_n \end{pmatrix}$

M(X) este cuprinsă între cea mai mică şi cea mai mare dintre valorile posibile ale variabilei.

Demonstrație.

Fie $a \le x_1, x_2, \ldots, x_n \le A$

$M(X) = p_1x_1 + p_2x_2 + \ldots + p_nx_n \rangle p_1a + \ldots + p_na = a$

$M(X) \langle p_1A + \ldots + p_nA = A$

$a \langle M(x) \langle A$

Alte valori tipice ale variabilei aleatoare

- *Momente de ordinu K* pentru variabila aleatoare *X*

Definiție. $M_k(X) = M(X^k)$

Dacă $X\begin{pmatrix} x_1 \qquad x_2 \ldots \quad x_n \\ p_1 \qquad p_2 \cdots \quad p_n \end{pmatrix}$

$$M_k(X) = p_1x_1^k + p_2x_2^k + \ldots + p_nx_n^k = \sum_{i=1}^{n} p_ix_i^k$$

- *Abaterea de la constanta* α *a variabilei X* este variabila $X - \alpha$

$M_k(X-\alpha) = M\left[(X-\alpha)^k\right]$

Dacă $\alpha = M(X) \Rightarrow M_k(X-\alpha)$ def *momentul centrat de ordinul K al variabilei X*

Definiție.: $X - M(X) =$ *abaterea de la medie a variabilei X* sau *abaterea variabilei X*

Scanned with OKEN Scanner

Observație.: $M(X - M(x)) = 0$

- Trebuie stabilit un indicator numeric al împrăştierii valorilor variabilei X în jurul valorii medii.

Împrăştierea variabilei X se caracterizează prin *valoarea medie a abaterii absolute*

$$|X - M(X)|$$

Definiție.: $M(|X - M(X)|) = \textit{abatere medie}$

- Dacă $X\begin{pmatrix} x_1 & x_2 \dots & x_n \\ p_1 & p_2 \dots & p_n \end{pmatrix}$ atunci *distribuția abaterii absolute* este

$|X - M(X)|\begin{pmatrix} |x_1 - m| & |x_2 - m| & |x_n - m| \\ p_1 & p_2 & p_n \end{pmatrix}$, iar abaterea medie

$$M(|X - M(X)|) = p_1|x_1 - m| + \dots + p_n|x_n - m| \quad m = M(X)$$

- Se poate lua ca indicator al împrăştierii variabile X şi

$$M[(X - m)^2] \text{ sau } M[(X - m)^4].$$

Observație. Folosirea *abaterii medii* $(M(|x - M(X)|)$ în calcule este incomodă. Foarte comodă este în schimb mărimea $M[(X - m)^2]$ numită dispersia variabilei aleatoare X.

Definiție.: Numim *dispersie a variabilei aleatoare X* momentul centrat de ordinul al doilea al acestei variabile.

$$\sigma^2 = D^2(X) = M[(X - m)^2], \quad m = M(X)$$

$$\Rightarrow D^2(X) = M(X^2) - [N(X)]^2$$

Proprietăți ale dispersiei

- Dispersia unei constante este nulă : $D^2(a) = 0$
- Fie două variabile aleatoare care diferă printr-o constantă :

  X, Y=X+a

Avem : $D^2(X) = D^2(Y)$

- $D^2(aX) = a^2D^2(X)$
- Fie $x_1, x_2, \dots, x_n$ variabile aleatoare independente.

$$D^2(X_1 + X_2 + \dots + X_n) = D^2(X_1) + D^2(X_2) + \dots + D^2(X_n)$$

Scanned with OKEN Scanner

Observație. De obicei, gradul de împrăştiere al unei variabile aleatoare X se exprimă nu prin dispersie, ci prin *abaterea medie pătratică D(X).*

$$\sigma = D(X) = \sqrt{D^2(X)} = \sqrt{M(X^2) - [M(X)]^2}$$

Definiție. *Abaterea medie pătratică* $\sigma = D(X) = \sqrt{D^2(X)}$

Proprietățile ale abaterii medii pătratice :

- D(a)=0, a=constantă
- D(X-a)=D(X)
- D(aX)=aD(X)
- Dacă $x_1, x_2, ..., x_n$ sunt variabile aleatorii independente

$$\sigma = D(X_1 + X_2 + ... + X_n) = \sqrt{\sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2}$$

Scanned with OKEN Scanner

3.4. Programe

3.4.1 COMPARATIE INTRE HISTOGRAMA SERIEI SI DISTRIBUTIA NORMALA

$i := 0..9$

$$a := \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 8 \\ 4 & 15 \\ 5 & 25 \\ 6 & 18 \\ 7 & 13 \\ 8 & 7 \\ 9 & 5 \\ 10 & 2 \end{pmatrix}$$

$a^{\langle 0 \rangle} =$

	0
0	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10

$a^{\langle 1 \rangle} =$

	0
0	3
1	4
2	8
3	15
4	25
5	18
6	13
7	7
8	5
9	2

$$m := \frac{\sum_{i=0}^{9} \left(a^{\langle 0 \rangle}\right)_i \cdot \left(a^{\langle 1 \rangle}\right)_i}{\sum_{i=0}^{9} \left(a^{\langle 1 \rangle}\right)_i} \qquad m = 5.4 \qquad n := \sum_{i=0}^{9} \left(a^{\langle 1 \rangle}\right)_i \qquad n = 10$$

$$D := \frac{\sum_{i=0}^{9} \left[\left(a^{\langle 0 \rangle}\right)_i - m\right]^2 \cdot \left(a^{\langle 1 \rangle}\right)_i}{\sum_{i=0}^{9} \left(a^{\langle 1 \rangle}\right)_i} \cdot \frac{n}{n-1} \qquad s := \sqrt{D} \qquad s = 1.954$$

$k := 0..2 \quad v_k := \left(a^{\langle 0 \rangle}\right)_0 \qquad k := 3..6 \quad v_k := \left(a^{\langle 0 \rangle}\right)_1 \qquad k := 7..14 \quad v_k := \left(a^{\langle 0 \rangle}\right)_2$

$k := 15..29 \quad v_k := \left(a^{\langle 0 \rangle}\right)_3 \qquad k := 30..54 \quad v_k := \left(a^{\langle 0 \rangle}\right)_4 \qquad k := 55..72 \quad v_k := \left(a^{\langle 0 \rangle}\right)_5$

$k := 73..85 \quad v_k := \left(a^{\langle 0 \rangle}\right)_6 \qquad k := 86..92 \quad v_k := \left(a^{\langle 0 \rangle}\right)_7 \qquad k := 93..97 \quad v_k := \left(a^{\langle 0 \rangle}\right)_8$

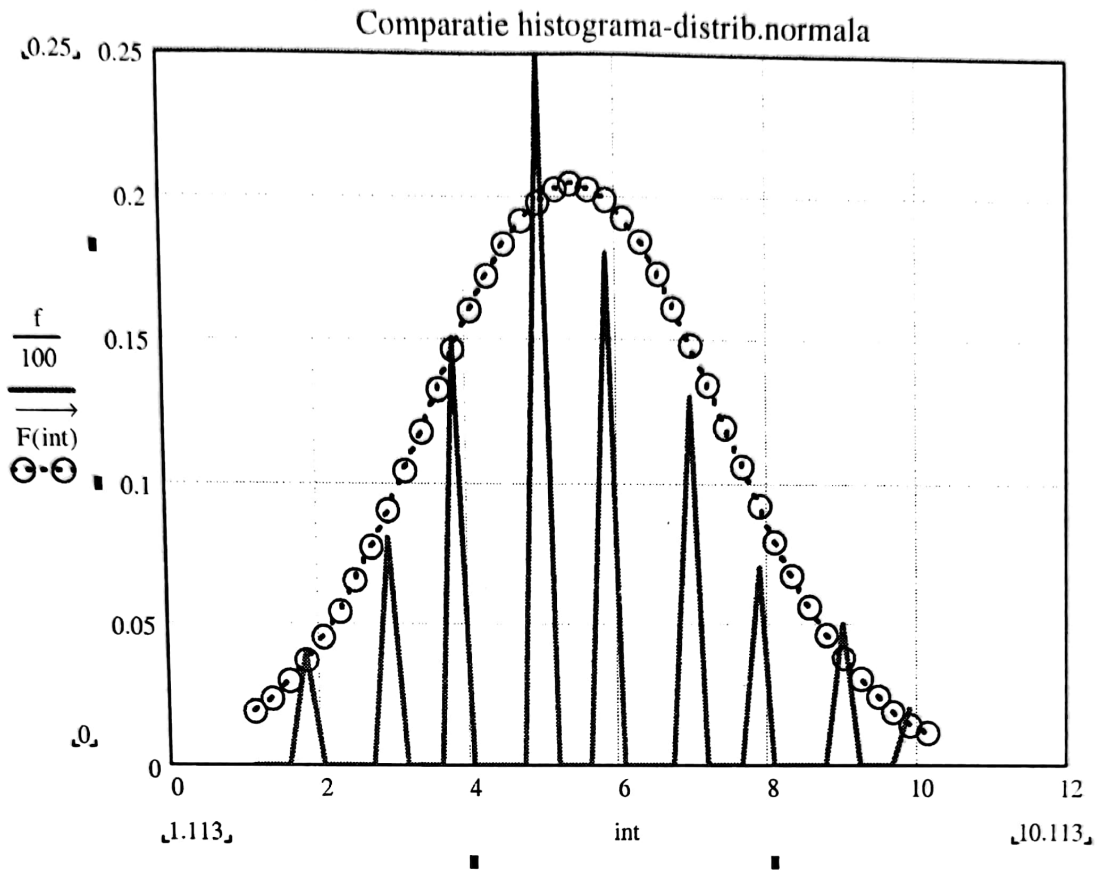
$k := 98..99 \quad v_k := \left(a^{\langle 0 \rangle}\right)_9 \qquad b := 40$

$l := \text{floor}\left(\min\left(a^{\langle 0 \rangle}\right)\right) \qquad u := \text{ceil}\left(\max\left(a^{\langle 0 \rangle}\right)\right) \qquad h := \frac{u - l}{b} \qquad l = 1 \quad u = 10 \quad h = 0.225$

$j := 0..b \qquad int_j := l + h \cdot j \qquad int := int + 0.5 \cdot h \qquad f := hist(int, v)$

$F(x) := dnorm(x, m, s) \qquad m_l := mean(v) \qquad m_l = 5.4 \qquad s_l := stdev(v) \cdot \sqrt{\frac{n}{n-1}} \qquad s_l = 1.954$

Scanned with OKEN Scanner



Scanned with OKEN Scanner

3.4.2. R E P A R T I T I A N O R M A L A . Densitatea de repartitie. Functia de repartitie

m=media **σ=abaterea standard**

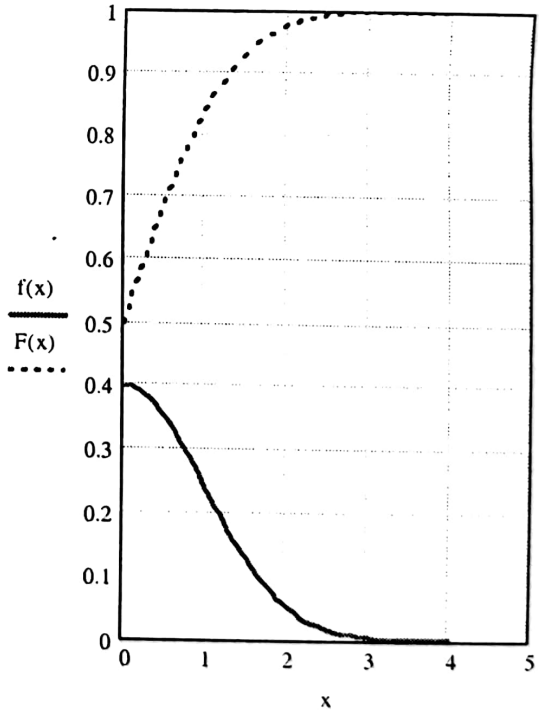
Densitatea de repartitie: $f(x) = \frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\exp\left[\frac{-1}{2}\cdot\left(\frac{x-m}{\sigma}\right)^2\right]$

Functia de repartitie: $F(x) = \int_{-\infty}^{x} f(t)\,dt$

$x := 0, 0.1 .. 4$ $\quad m := 0$ $\quad \sigma := 1$

$$f(x) := \frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\exp\left[\frac{-1}{2}\cdot\left(\frac{x-m}{\sigma}\right)^2\right] \qquad F(x) := \int_{-\infty}^{x} \frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\exp\left[\frac{-1}{2}\cdot\left(\frac{t-m}{\sigma}\right)^2\right] dt$$

x =	f(x) =	F(x) =
0	0.399	0.5
0.1	0.397	0.54
0.2	0.391	0.579
0.3	0.381	0.618
0.4	0.368	0.655
0.5	0.352	0.691
0.6	0.333	0.726
0.7	0.312	0.758
0.8	0.29	0.788
0.9	0.266	0.816
1	0.242	0.841
1.1	0.218	0.864
1.2	0.194	0.885
1.3	0.171	0.903
1.4	0.15	0.919
1.5	0.13	0.933



Scanned with OKEN Scanner

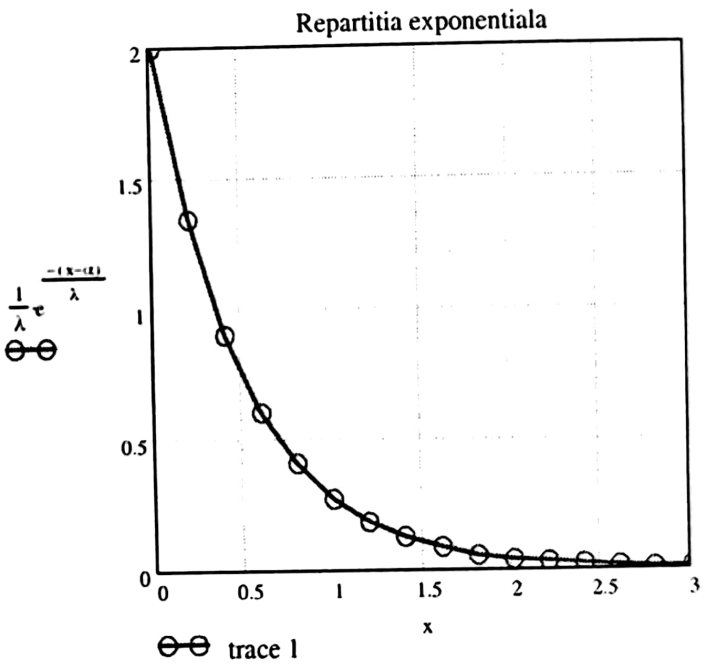
3.4.3. REPARTITIA EXPONENTIALA E (λ, α)

Densitatea de repartitie

$$f_X(x) = \begin{cases} \frac{1}{\lambda}\cdot e^{\frac{-(x-\alpha)}{\lambda}} \leftarrow x \geq \alpha, \lambda > 0 \\ 0 \leftarrow x < 0 \end{cases} \qquad x := 0, 0.2 .. 3 \qquad \alpha := 0 \quad \lambda := 0.5$$

$$f(x) := \frac{1}{\lambda}\cdot e^{\frac{-(x-\alpha)}{\lambda}} \qquad F(x) := \int_0^x \frac{1}{\lambda}\cdot e^{\frac{-(x-\alpha)}{\lambda}}\, dx$$

x =	f(x) =
0	2
0.2	1.341
0.4	0.899
0.6	0.602
0.8	0.404
1	0.271
1.2	0.181
1.4	0.122
1.6	0.082
1.8	0.055
2	0.037
2.2	0.025
2.4	0.016
2.6	0.011
2.8	$7.396 \cdot 10^{-3}$
3	$4.958 \cdot 10^{-3}$



3.4.4. REPARTITIA STUDENT(t)

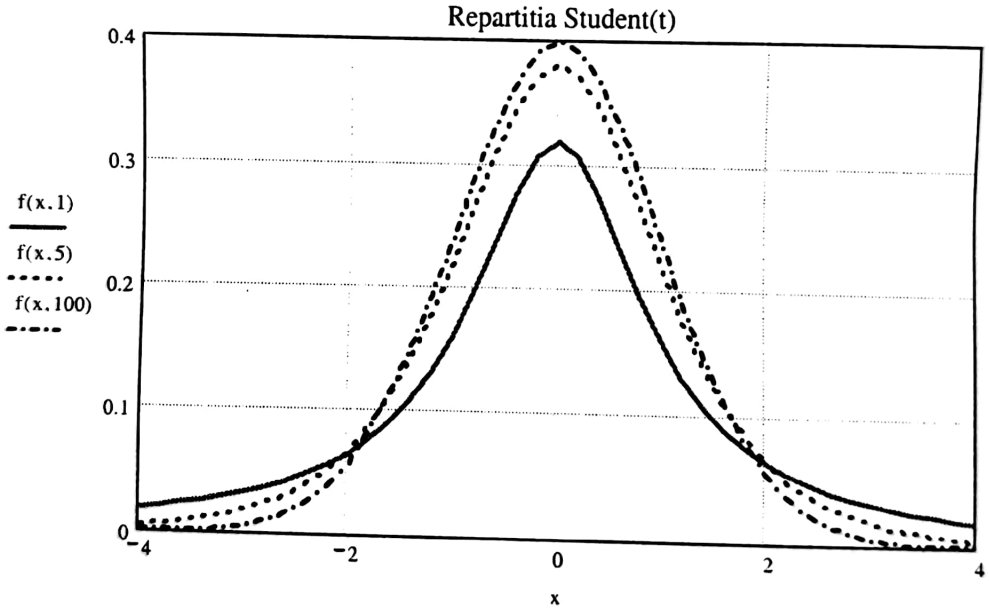
Densitatea de repartitie

$$f_X(x) \equiv \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{\pi\cdot n}}\cdot\left(1+\frac{x^2}{n}\right)^{\frac{-(n+1)}{2}}$$

x ≡ real n=numar grade de libertate

$x := -4, -3.8 .. 4 \qquad n := 8$

$$f(x,n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{\pi\cdot n}}\cdot\left(1+\frac{x^2}{n}\right)^{\frac{-(n+1)}{2}}$$



$$\int f(x,1)\,dx \rightarrow \frac{\operatorname{atan}(x)}{\pi}$$

Scanned with OKEN Scanner

3.4.5. REPARTITIA HI-PATRAT $\chi 2$

Densitatea de repartitie

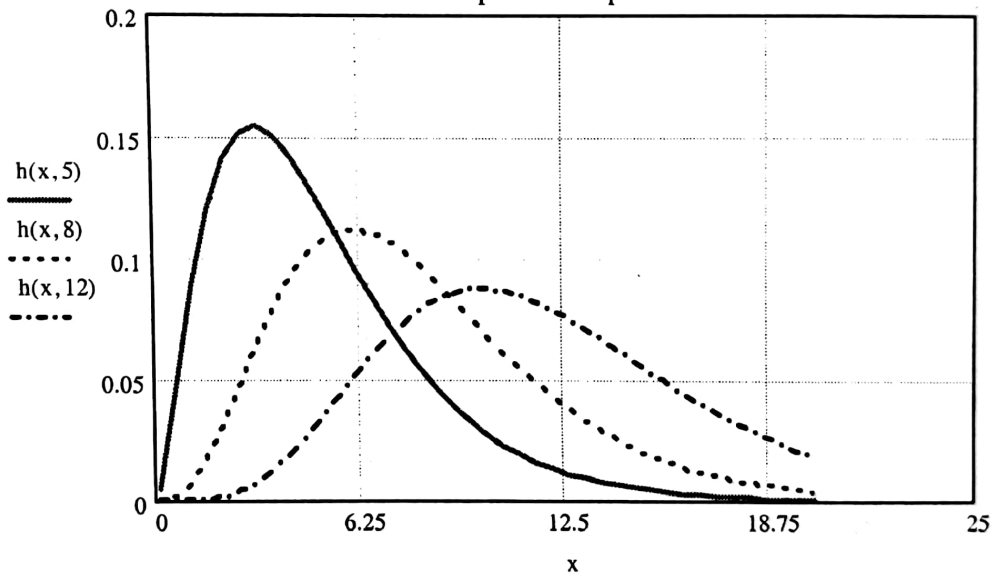
$$h_X(x,n) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\cdot\Gamma\left(\frac{n}{2}\right)}\cdot x^{\frac{n}{2}-1}\cdot e^{\frac{-x}{2}} & \leftarrow x>0 \\ 0 & \leftarrow x\le 0 \end{cases}$$

$$h(x,n) := \frac{1}{2^{\frac{n}{2}}\cdot\Gamma\left(\frac{n}{2}\right)}\cdot x^{\frac{n}{2}-1}\cdot e^{\frac{-x}{2}}$$

n=numarul gradelor de libertate

$x := 0.1, 0.6 .. 20.1$

Repartitia hi-patrat



$$\int h(x,n)\,dx \rightarrow \frac{-1}{48}\cdot x^3\cdot\exp\left(\frac{-1}{2}\cdot x\right) - \frac{1}{8}\cdot x^2\cdot\exp\left(\frac{-1}{2}\cdot x\right) - \frac{1}{2}\cdot x\cdot\exp\left(\frac{-1}{2}\cdot x\right) - \exp\left(\frac{-1}{2}\cdot x\right)$$

$n := 1$ $\quad \int h(x,n)\,dx \rightarrow \operatorname{erf}\left[\frac{1}{2}\cdot\sqrt{2}\cdot x^{\left(\frac{1}{2}\right)}\right]$ $\quad \operatorname{erf}\left[\frac{1}{2}\cdot\sqrt{2}\cdot 0^{\left(\frac{1}{2}\right)}\right] = 0$

$n := 2$ $\quad \int h(x,n)\,dx \rightarrow -\exp\left(\frac{-1}{2}\cdot x\right)$ $\quad -\exp\left(\frac{-1}{2}\cdot 0\right) = -1$

$n := 6$ $\quad \int h(x,n)\,dx \rightarrow \frac{-1}{8}\cdot x^2\cdot\exp\left(\frac{-1}{2}\cdot x\right) - \frac{1}{2}\cdot x\cdot\exp\left(\frac{-1}{2}\cdot x\right) - \exp\left(\frac{-1}{2}\cdot x\right)$

Scanned with OKEN Scanner

$$\lim_{x\to 0} \frac{-1}{8}\cdot x^2\cdot\exp\left(\frac{-1}{2}\cdot x\right) - \frac{1}{2}\cdot x\cdot\exp\left(\frac{-1}{2}\cdot x\right) - \exp\left(\frac{-1}{2}\cdot x\right) \to -1$$

$n := 1 \qquad F(\chi^2, n) = 0.95 \qquad \operatorname{erf}\left[\frac{1}{2}\cdot\sqrt{2}\cdot(y)^{\left(\frac{1}{2}\right)}\right] - 0.95 = 0$

$y := 5$

$$sol := root\left[\operatorname{erf}\left[\frac{1}{2}\cdot\sqrt{2}\cdot(y)^{\left(\frac{1}{2}\right)}\right] - 0.95, y\right] \qquad sol = 3.874$$

$y := 3.874 \qquad sol := root\left[\operatorname{erf}\left[\frac{1}{2}\cdot\sqrt{2}\cdot(y)^{\left(\frac{1}{2}\right)}\right] - 0.95, y\right] \qquad sol = 3.841 \qquad (\chi^2)_{0.95} = 3.84$

3.4.6. REPARTITIA F

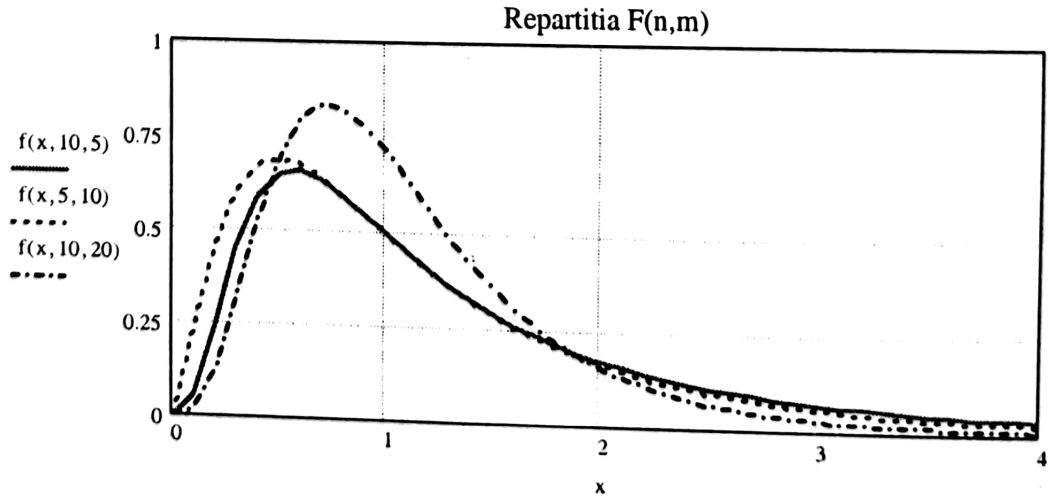
Repartitia F notata F(n,m) are n,respectiv m grade de libertate.

Densitatea de repartitie:

$$f_X(x,n,m) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)}\cdot\left(\frac{n}{m}\right)^{\frac{n}{2}}\cdot x^{\frac{n}{2}-1}\cdot\left(1+\frac{n}{m}\cdot x\right)^{\frac{-(n+m)}{2}} \qquad x \geq 0$$

$x := 0, 0.1 .. 4$

$$f(x,n,m) := \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)}\cdot\left(\frac{n}{m}\right)^{\frac{n}{2}}\cdot x^{\frac{n}{2}-1}\cdot\left(1+\frac{n}{m}\cdot x\right)^{\frac{-(n+m)}{2}}$$



Scanned with OKEN Scanner

$x := 0, 0.3 .. 3$

$x =$	$f(x, 10, 20) =$	$\int_0^x f(x, 10, 20)\, dx =$
0	0	0
0.3	0.311	0.027
0.6	0.792	0.204
0.9	0.779	0.45
1.2	0.563	0.652
1.5	0.358	0.789
1.8	0.216	0.874
2.1	0.128	0.924
2.4	0.076	0.954
2.7	0.045	0.972
3	0.027	0.982

$$\int f(x, 1, 2)\, dx \rightarrow \frac{1}{(4 + 2 \cdot x)^{\left(\frac{1}{2}\right)}} \cdot x^{\left(\frac{1}{2}\right)} \cdot \sqrt{2}$$

$$\lim_{x \to 0} \frac{1}{(4 + 2 \cdot x)^{\left(\frac{1}{2}\right)}} \cdot x^{\left(\frac{1}{2}\right)} \cdot \sqrt{2} \rightarrow 0$$

$x := 5$ $\quad sol := root\left[\frac{1}{(4 + 2 \cdot x)^{\left(\frac{1}{2}\right)}} \cdot x^{\left(\frac{1}{2}\right)} \cdot \sqrt{2} - 0.95, x\right]$ $\quad sol = 18.386$

$x := 18.386$ $\quad sol := root\left[\frac{1}{(4 + 2 \cdot x)^{\left(\frac{1}{2}\right)}} \cdot x^{\left(\frac{1}{2}\right)} \cdot \sqrt{2} - 0.95, x\right]$ $\quad sol = 18.512$

Scanned with OKEN Scanner

CAPITOLUL 4

SELECȚIE. TEORIA ELEMENTARĂ A SONDAJULUI

4.1. Noțiuni generale. Indicatori.

În cazul cantitativ, o caracteristică se numeşte variabilă şi poate fi discretă sau continuă.

Problema fundamentală a statisticii matematice este următoarea: dacă nu cunoaştem şi nici nu putem determina *apriori legea de distribuție* (sau de repartiție) a unei populații statistice pe care o studiem (după una sau mai multe caracteristici), cum putem găsi, totuşi, pe calea experiențelor pe care le facem asupra populației statistice, această *lege de repartiție* sau măcar unele caracteristici principale (sau informații *care să descrie evoluția populației studiate* în vederea unei previziuni a evoluției viitoare?

- *În anumite condiții*, acest lucru este posibil, *iar procedeul de a obține anumite informații* care se prelucrează şi analizează ulterior reprezintă *metoda selecției.*

Definiție. Se numeşte selecție orice alegere, la întâmplare sau după anumite reguli, a unor elemente ale populației statistice pentru a fi studiate cu scopul de a obține informații asupra întregii populații.

Definiție. Mulțimea elementelor alese prin selecție se numeşte *eşantion*, iar numărul elementelor selecționate se numeşte *volum al eşantionului* (sau volum al selecției

Observație. Când este vorba de oameni, dar nu numai atunci, se foloseşte termenul de *sondaj* în locul celui de selecție, numărul persoanelor chestionate alcătuind *volumul sondajului.*

- Fie X — o caracteristică a unei populații statistice. Vom considera că X este o variabilă aleatoare *a cărei repartiție ne interesează* şi pe care o numim *repartiție teoretică.*

Observație. Repartiția teoretică poate fi dată de *funcția* de repartiție teoretică $F(x)$, de densitatea de repartiție teoretică $f(x)$ în cazul continuu sau de probabilitățile teoretice $p(x)=P(X=x)$ în cazul discret.

Scanned with OKEN Scanner

De regulă, repartiția depinde și de un parametru θ care poate fi :

— scalar ;

— vector real ;

astfel că scriem $F(x, \theta)$, $f(x, \theta)$, $p(x, \theta)$.

- Dacă nu se cunoaşt θ, atunci se spune că *repartiția* variabilei aleatoare X *este specificată*.
- Dacă, în plus, se cunoaşte şi valoarea parametrului θ, atunci repartiția variabilei aleatoare X este *complet specificată*.

Observație. Se urmărește ca pe baza sondajului (sau selecției) de volum dat să se deducă repartiția probabilistică a caracteristicii aleatoare X asimilată printr-o variabilă aleatoare.

Indicatori numerici

Pe baza sondajului de volum n asupra caracteristicii X şi ale

cărui rezultate sunt numerele x_i cu frecvențele absolute n_i, $1 \le i \le m$.

1. *Media aritmetică* $(\bar{x})$

Proprietăți ale mediei aritmetice

- $\min_i \{x_i\} \le \bar{x} \le \max_i \{x_i\}$
- $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$
- $(\forall) a = const : \sum_{i=1}^{n} (x_i - a)^2 \ge \sum_{i=1}^{n} (x_i - \bar{x})^2$

(suma pătratelor abaterilor are valoarea cea mai mică atunci când abaterea se calculează față de $\bar{x}$)

1. MEDIA GEOMETRICĂ (G)
2. MEDIA ARMONICĂ (H)
3. MEDIANA
4. MODUL SAU MODA (M_o)

Scanned with OKEN Scanner

Indicatori numerici ai variației observațiilor

1. AMPLITUDINEA
2. ABATEREA MEDIE ABSOLUTĂ (E)
3. DISPERSIA DE SELECȚIE. ABATEREA MEDIE PĂTRATICĂ

Notată cu $\underbrace{S^2}$, *dispersia de selecție* are expresia:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - (\bar{x})^2$$

în cazul valorilor distincte sau expresia:

$$S^2 = \frac{1}{n}\sum_{i=1}^{m}n_i(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{m}n_i x_i^2 - (\bar{x})^2$$

pentru valori care se repetă.

Adeseori se foloseşte *abaterea medie pătratică modificată* $\underbrace{\tilde{S}^2}$

având expresia:

$$\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}S^2$$

Definiție. Numărul S s $\tilde{S}$ se numeşte *abaterea medie pătratică de selecție.*

Observație. Dispersia de selecți S^2 are aceeaşi semnificație ca şi dispersia teoretică, şi anume aceea de *măsură* a gradului de împrăştiere (sau de dispersare) a observațiilor în jurul valorii medii a acestora $\bar{x}$.

*Observație. C*a şi în cazul teoretic când s-au prezentat *momentele inițiale şi momentele centrate ale unei variabile aleatoare*, le putem introduce şi pentru observațiile $\{x_i\}$ asupra caracteristicii aleatoare X.

- *Momentele inițiale de selecție de ordinul* K $(\overline{M}_k)$

$$\overline{M}_k = \frac{1}{n}\left(x_1^k + x_2^k + \ldots + x_n^k\right) = \frac{1}{n}\sum_{i=1}^{n}x_i^k$$

în cazul observaţiilor distincte, sau

$$\overline{M}_k = \frac{1}{n}\left(n_1x_1^k + n_2x_2^k + \ldots + n_mx_m^k\right) = \frac{1}{n}\sum_{i=1}^{m}n_ix_i^k$$

atunci când există *m* observații distincte din cele *n* efectuate.

Scanned with OKEN Scanner

- **Momente centrate de selecție de ordinul K ($\bar{\mu}_K$)**

$$\bar{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k \quad \text{în cazul observațiilor distincte} \quad \text{sau}$$

$$\bar{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \bar{x})^k \quad \text{când valorile se repetă.}$$

Observație. — Dacă k =1, atunci $\bar{M}_1 = \bar{x}$ și $\mu_1 = 0$.

— Dacă (Pentru) k =2, avem: $\bar{\mu}_2 = S^2 = \bar{M}_2 - (\bar{M}_1)^2$.

- **Funcția de repartiție de selecție**

O prezentare riguroasă a conceptului de *selecție*, din care rezultă importanța sa fundamentală în statistica matematică, se realizează cu ajutorul teoriei probabilităților.

Fie un experiment aleator căruia îi asociem o variabilă aleatoare *X*.

Să presupunem că experimentul se repetă de n ori în mod independent, sau că asupra lui *X* se fac şi probe (experienţe) independente. Asociem fiecărei *probe i* o variabilă aleatoare $x_i (i = \overline{1,n})$ suntem conduşi la *n* variabile independente $x_1, x_2, ..., x_n$ sau la un *vector* aleator *n*-dimensional $(x_1, x_2, ..., x_n)$ ale cărei componente sunt variabile aleatoare independente în totalitate.

- Dacă *X* are funcția de repartiție *F(X)*, atunci *funcțiile de repartiție individuale*

$$F_i(x_i) = F(x_i) \quad i = \overline{1,n}$$

şi funcția de repartiție comună :

$$F(x_1, x_2, ..., x_n) = F(x_1) \cdot F(x_2) ... F(x_n)$$

Definiție. Se numeşte selecție de volum n asupra lui *X* şi se notează $(x_1, x_2, ..., x_n)$ o realizare a vectorului aleator $(X_1, X_2, ..., X_n)$ (cu $X_i = x_i$ – adică la experienţa *i* asupra lui *X* s-a obținut rezultatul x_i).

Definiție. Variabila aleatoare *X(n)* asociată experimentului cu n probe independente efectuate asupra lui *X* şi definită astfel :

Scanned with OKEN Scanner

$$X_{(n)}:\begin{bmatrix} x_1 & x_2 \dots & x_n \\ \frac{1}{n} & \frac{1}{n} \dots & \frac{1}{n} \end{bmatrix}$$ se numeşte *variabilă aleatoare empirică* sau *variabilă aleatoare de selecţie*

Observaţie. Presupunem că valorile variabilei de selecţie sunt ordonate strict crescător :

$$X_{(n)}:\begin{bmatrix} x_1 & x_2 \dots & x_m \\ \frac{n_1}{n} & \frac{n_2}{n} \dots & \frac{n_m}{n} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \dots & x_m \\ p_1 & p_2 \dots & p_m \end{bmatrix}$$

unde : $n_1 + n_2 + \dots + n_m = n$, n_i — frecvenţe absolute, iar

$p_i = \frac{n_i}{n} (i = \overline{1,m})$ dintre variabilele de selecţie şi variabilele aleatoare teoretice discrete, dacă avem în vedere şi legea numerelor mari.)

Aplicaţie. Se consideră la întâmplare 10 cereri de creditare depuse la o bancă într-o anumită perioadă de timp.

Numărul cererii	1	2	3	4	5	6	7	8	9	10
Suma solicitată (mil.u.m.)	5	3	5	1	4	3	2	5	5	2

Variabila de selecţie X(20) este următoarea :

$$X(10):\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ \frac{1}{10} & \frac{2}{10} & \frac{2}{10} & \frac{1}{10} & \frac{4}{10} \end{bmatrix}$$ ← valoarea variabilei de selecţie ← frecvenţe relative

Definiţie. Funcţie empirică de repartiţie

$$F_n(x) = \sum_{x_i < x} p_i \begin{cases} 0 & x \le x_1 \\ p_1 + p_2 + \dots + p_k & x_k < x \le x_{k+1} \quad 1 \le k \le n-1 \\ 1 & x > x_n \end{cases}$$

sau, sub formă detaliată :

Scanned with OKEN Scanner

$$F_n(x)=\begin{cases}0 & x\le x_1\\ p_1 & x_1<x\le x_2\\ p_1+p_2 & x_2<x\le x_3\\ \dots\dots\dots\

Fie *Fn(x)* şi *F(x)* funcţii ataşate unei populaţii statistice, *F(x)* — continuă. Avem:

$$\lim_{n\to\infty} P\left(Dn\rangle z\sqrt{n}\right)=\begin{cases}2\sum_{k=1}^{\infty}(-1)^{k-1}\exp\left(-2k^2z^2\right), & z\rangle 0\\ 0 & z\langle 0\end{cases}$$

unde $Dn=\sup\left|Fn(x)-F(x)\right|$

Observaţie. Cele două teoreme constituie suportul teoretic al convergenţei repartiţiei empirice către repartiţia teoretică a variabilei aleatoare *X*. *Se poate spune că în urma unei selecţii de volum suficient de mare* variabila de selecţie *X(n)* descrie suficient de bine variabila aleatoare teoretică pe care, de fapt, o căutăm.

Momente de selecţie

Definiţie. Fiind date variabilele aleatoare independente şi identic repartizate $X_1, X_2, ..., X_n$, care definesc o selecţie de volum n asupra caracteristicii aleatoare X a unei populaţii statistice, atunci orice funcţie $Tn(X_1, X_2, ..., X_n)$ se numeşte funcţie de selecţie sau statistică.

Ele se definesc în continuare.

Definiţie. Se numeşte *moment iniţial de selecţie de ordinul r* statistica :

$$\overline{M}_r(X)=\frac{1}{n}\sum_{i=1}^{n}X_i^r=\overline{M}_r$$

- *Dacă r=1, momentul iniţial de selecţie de ordinul unu* se numeşte *medie de selecţie* $(\overline{X})$.

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

Observaţie. Este necesară distincţia clară dintre statistica $\overline{X}$ şi o valoare a sa, corespunzătoare unui sondaj anume, notată $\overline{x}$.

$\left(\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\right)\leftarrow$ denumită la fel *medie de selecţie*

Observaţie. Orice statistică $T(X_1, X_2, ..., X_n)$ este o variabilă aleatoare.

Scanned with OKEN Scanner

Propoziție. Dacă $X_1, X_2, ..., X_n$ este o selecție de volum n independentă asupra caracteristicii X a unei populații statistice, cu $M(X) = m$ și $D^2(X) = \sigma^2$, iar $\overline{X}$ este media de selecție corespunzătoare, atunci $M(\overline{X}) = m$; $D^2(\overline{X}) = \frac{\sigma^2}{n}$.

Demonstrație.

$$M(\overline{X}) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n}(n \cdot m) = m$$

$$D^2(\overline{X}) = D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Propoziție. Dacă $X_1, X_2, ..., X_n$ este o selecție independentă de volum n asupra caracteristicii aleatoare X care are momente teoretice de ordinul *r* date de relația: $Mr(X) = M(X^r) = m_r$ iar $\overline{Mr}(X)$ este momentul inițial de selecție de ordinul *r* corespunzător, atunci :

$$M\left[\overline{Mr}(X)\right] = m_r; \qquad D^2\left[\overline{Mr}(X)\right] = \frac{1}{n}(m_{2r} - m_r^2)$$

Demonstrație.

$$M\left[\overline{Mr}(X)\right] = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i^r\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i^r) = \frac{1}{n} n m_r = m_r$$

$$D^2\left[\overline{Mr}(X)\right] = D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i^r\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} D^2(X_i^r)\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} (X_i^r - m_r)^2\right] =$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i^{2r} - 2m_r\sum_{i=1}^{n} X_i^r + nm_r^2\right) = \frac{1}{n^2}(nm_{2r} - 2nm_r^2 + nm_r^2) =$$

$$= \frac{1}{n}(m_{2r} - m_r^2)$$

Definiție. Moment centrat de selecție de ordinul *r* este statistica :

$$\overline{\mu}_r(X) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^r = \overline{\mu}_r$$

Scanned with OKEN Scanner

Dacă r=2, avem $\bar{\mu}_2(X) = S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ care pentru o selecție $(x_1, x_2, ..., x_n)$ devine un număr s^2, adică :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

Se poate introduce:

- *DISPERSIA de selecție modificată*

$$\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}S^2$$

căreia îi corespunde valoarea numerică :

$$\tilde{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}s^2$$

- *Relație între momentele centrate de selecție* $\bar{\mu}_r$ *și momentele inițiale de selecție* $\bar{M}_r$

$$\bar{\mu}_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^r =$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}X_i^r - C_r^1\bar{X}\sum_{i=1}^{n}X_i^{r-1} + C_r^2X^2\sum_{i=1}^{n}X_i^{r-2} - ... + (-1)^r\bar{X}^r \cdot n\right)$$

Dar : $\frac{1}{n}\sum_{i=1}^{n}X_i^{r-k} = \bar{M}_{r-k} \qquad 0 \le k \le r-1$

și avem :

$$\bar{\mu}_r = \bar{M}_r - C_r^1\bar{M}_1\bar{M}_{r-1} + C_r^2\bar{M}_1^2\bar{M}_{r-2} + ... + (-1)^r\bar{M}_1^r$$

- *Particularizări*

$$\bar{\mu}_1 = 0$$

$$\bar{\mu}_2 = \bar{M}_2 - \bar{M}_1 = S^2$$

$$\bar{\mu}_3 = \bar{M}_3 - 3\bar{M}_1\bar{M}_2 + 2\bar{M}_1^3$$

Propoziție. Dacă $x_1, x_2, ..., x_n$ este o selecție independentă de volum n efectuată asupra unei populații statistice având caracteristica aleatoare

Scanned with OKEN Scanner

X, cu media teoretică $M(X) = m$, dispersia teoretică $D^2(X) = \sigma^2$ şi moment centrat teoretic μ_4, atunci pentru dispersia de selecţie $S^2\left[S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]$ şi dispersia de selecţie modificată $\tilde{S}^2\left[\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]$, avem :

— valorile medii

$$M(S^2) = \frac{n-1}{n}\sigma^2; \qquad M(\tilde{S}^2) = \sigma^2$$

şi — dispersiile :

$$D^2(S^2) = \frac{(n-1)^2\mu_4 - (n-1)(n-3)\sigma^4}{n^3}; \quad D^2(S^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right)$$

(Fără demonstraţie)

Observaţie.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - m) + (m - \overline{X})\right]^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - m) + (m - \frac{1}{n}\sum_{i=1}^{n}X_i)\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - m) - \frac{1}{n}\sum_{i=1}^{n}X_i - m)\right]^2$$

- $M\left[(X_i - m)^2\right] = \sigma^2$
- Dacă $Y = S^2$

$$D^2(Y) = D^2(S^2) = M(Y^2) - M^2(Y) = M(S^4) - M^2(S^2)$$

- Când volumul de selecţie este foarte mare ($n \to \infty$)

$$\lim_{n\to\infty} D^2(S^2) = 0 = \lim_{n\to\infty} D^2(\tilde{S}^2)$$

Selecţii din populaţii normale

Propoziţie. Fie $x_1, x_2, \ldots, x_n$ o selecţie independentă de volum n efectuată asupra unei populaţii statistice având caracteristica aleatoare X, pentru

Scanned with OKEN Scanner

care există media M(X)=m, dispersia $D^2(X) = M\left[(x-m)^2\right] = \sigma^2 \neq 0$, precum şi momentul $H^3 = M\left[|x-m|^3\right]$.

În aceste condiţii, *variabila aleatoare redusă* $Y = \dfrac{\overline{X} - m}{\dfrac{\sigma}{\sqrt{n}}}$ urmează o repartiţie asimptotică N (0,1).

Demonstraţie.

Notăm cu $Z_i = \dfrac{X_i}{n}$. Avem $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \dfrac{X_i}{n} = \sum_{i=1}^{n} Z_i$

$$M(Z_i) = M\left(\frac{X_i}{n}\right) = \frac{1}{n} M(X_i) = \frac{m}{n}$$

$$D^2(Z_i) = M\left[\left(Z_i - \frac{m}{n}\right)^2\right] = M\left[\left(\frac{X_i - m}{n}\right)^2\right] = \frac{1}{n^2} M\left[(X_i - m)^2\right] = \frac{\sigma^2}{n^2}$$

$$H_i^3(Z_i) = M\left(\left|Z_i - \frac{m}{n}\right|^3\right) = M\left(\frac{|X_i - m|^3}{n^3}\right) = \frac{1}{n^3} M\left(|X_i - m|^3\right) = \frac{H^3}{n^3}$$

Obţinem:

$$S_n^2 = \sum_{i=1}^{n} D^2(Z_i) = n \cdot \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$K_n^3 = \sum_{i=1}^{n} H_i^3 = n \frac{H^3}{n^3} = \frac{H^3}{n^2}$$

Deoarece $\lim\limits_{n\to\infty} \dfrac{K_n}{S_n} = \lim\limits_{n\to\infty} \dfrac{H}{n^{2/3}} \cdot \dfrac{n^{\frac{1}{2}}}{\sigma} = \lim\limits_{n\to\infty} \dfrac{H}{\sigma} \cdot n^{-\frac{1}{6}} = 0$ rezultă, potrivit teoremei lui Liapunov că variabila aleatoare Y este repartizată N (0,1).

Scanned with OKEN Scanner

4.2. **Teoria elementară a sondajului.**

Metoda sondajului oferă posibilitatea dobândirii de informații referitoare la întreaga colectivitate prin observarea numai a unei părți a acesteia. Această parte este numită *eşantion* sau *selecție*, provenită din *populația originară* (sau *colectivitatea generală*) care indică întreaga populație luată în studiu.

Teoria sondajului este, de fapt, studiul relațiilor existente între o populație şi eşantionul extras din populația analizată.

Teoria este importantă în special în controlul calității produselor sau atunci când cercetarea întregii colectivități originare necesită costuri mari.

Cerința generală a unui sondaj

Rezultatul sondajului *să fie reprezentativ* pentru întreaga colectivitate. Se admite ca suficient de reprezentativ sondajul care nu conduce la erori față de colectivitatea originară mai mari de ±5%.

Condiții ce se impun eşantioanelor

- **Procedee de alcătuire a eşantionului**

Eşantioanele trebuie să fie reprezentative. Ele vor îndeplini următoarele condiții :

1. populația originară să fie cât mai omogenă ;
2. unitățile eşantionului să fie extrase la întâmplare ;
3. eşantionul să fie suficient de mare pentru ca în el să apară trăsăturile populației originare ;
4. independența unităților extrase între ele.

Alegerea unităților ce vor forma eşantionul se poate face :

a) pe baza unei scheme probabilistice ;

b) în mod dirijat.

Se utilizează trei procedee :

1. procedeul tragerii la sorți sau al loteriei ;
2. procedeul selecției bazate pe tabelul cu numere aleatoare ;
3. procedeul alcătuirii mecanice.

Scanned with OKEN Scanner

1. **Procedeul tragerii la sorți**

Se poate realiza sub două forme :

1.1. cu restituirea unității extrase (selecție întâmplătoare repetată) ;

1.2. fără restituirea unității extrase (selecție întâmplătoare nerepetată)

2. **Procedeul selecției bazate pe** *tabelul cu numere aleatoare.*

Tabelele cu numere aleatoare se întocmesc cu ajutorul mașinilor de amestecat numere. Numerele din tabel sunt așezate pe coloane și rânduri într-o ordine absolut întâmplătoare. Pentru formarea eșantionului, tot la întâmplare se stabilesc coloanele și rândurile pentru citirea numerelor întâmplătoare.

Dacă anumite numere din tabel depășesc pe cele din colectivitatea originară, acestea se exclud din selecție. Și în acest caz alegerea coloanelor și rândurilor de selecție se poate face repetat sau nerepetat.

3. **Procedeul alcătuirii mecanice a eșantionului**

Unitățile populației originare se înscriu, într-o anumită ordine, într-un tabel din care se aleg apoi unitățile aflate la distanțe egale. Se alege pasul de numărare $\left(p = \frac{N}{n} \right)$ unde N = nr. unităților populației originare; n = nr. unităților eșantionului.

Prima unitate se selecționează la întâmplare, a doua unitate selecționată este unitatea aflată la distanța de un pas față de prima unitate, ș.a.m.d.; unitatea a n – a din eșantion se va afla la distanța de (n – 1) pași față de prima unitate selecționată.

- **Sondajul simplu repetat și nerepetat**

Fie o urnă cu N bile, fiecare bilă având inscripționat un număr.

Se extrage din urnă o bilă (număr). Dacă numărul extras se va introduce în urnă, înaintea următoarei extrageri spunem că *sondajul este repetat*, iar în caz contrar, *sondaj nerepetat*.

Dacă extragem succesiv 5 bile în mod nerepetat, dintr-o urnă cu 50 bile, se formează un eșantion dintr-o populație finită. Dacă înaintea fiecărei extrageri bila se reântoarce în urnă spunem că populația este infinită.

- **Distribuții de selecție**

Să considerăm toate eșantioanele posibile de mărime „n" care pot fi extrase dintr-o populație dată (cu sau fără repetare).

Scanned with OKEN Scanner

Pentru fiecare eşantion se calculează un parametru, cum ar fi media aritmetică, abaterea pătratică medie, etc. care poate varia de la un eşantion la altul. *În acest mod obţinem o distribuţie* a parametrilor eşantioanelor.

Exemplu : distribuţia mediei eşantionului sau distribuţia abaterii mediei pătratice, a dispersiei, medianei pentru colectivitatea de selecţie.

- **Pentru orice distribuţie de selecţie** se poate calcula media, abaterea pătratică medie, dispersia etc.

Distribuţia mediei eşantionului

Presupunem că sunt extrase nerepetat toate eşantioanele de mărime „n" dintr-o colectivitate originară finită de mărime N (N > n).

Notăm $\bar{x}_s$ şi ∇_s – media şi abaterea pătratică medie a distribuţiei eşantionului.

$\bar{x}_0$ şi ∇_0 – media şi abaterea pătratică medie a colectivităţii generale

$$\bar{x}_s = x_0$$

şi avem :

$$\nabla_s = \frac{\nabla_0}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad (4.1)$$

Dacă populaţia originară este infinită ($N \to \infty$) sau dacă sondajul este repetat, relaţiile (4.1) devin

$$\bar{x}_s = \bar{x}_0$$

$$\nabla_s = \frac{\nabla_0}{\sqrt{n}} \qquad (4.2)$$

Pentru valori mari ale lui „n" ($n \geq 30$) distribuţia mediei eşantionului este o distribuţie normală cu media $\bar{x}_s$ şi abaterea pătratică medie ∇_s. În cazul în care populaţia are o distribuţie normală şi distribuţia mediei eşantionului este normală chiar pentru valori mici ale lui n ($n < 30$).

Scanned with OKEN Scanner

Tipuri de sondaj

Deosebim patru tipuri de sondaj :

1. sondaj simplu
2. sondaj mecanic
3. sondaj tipic (stratificat)
4. sondaj de serii

Fiecare tip de sondaj poate fi repetat şi nerepetat.

Observaţie: *Pentru toate tipurile de sondaj se calculează trei* indicatori:

1. *eroare medie de sondaj* (μ) (sau eroare medie de reprezentativitate)
2. *eroare limită admisă*, notată cu
   Δx — pentru caracteristici nealternative
   Δw — pentru caracteristici alternative
3. *volumul eşantionului* pentru caracteristici alternative şi nealternative.

- **Sondajul simplu repetat**

1. Eroarea medie de sondaj

a. *pentru caracteristici nealternative*

$$\nabla_0^2 = \nabla_s^2 \frac{n}{n-1} \underset{(\mu)}{\Longrightarrow} n\mu^2 = \nabla_s^2 \frac{n}{n-1};\ \mu^2 = \frac{\nabla_s^2}{n-1} \Rightarrow$$

$$\nabla_0^2 \doteq n \cdot \mu^2$$

$$\mu^2 \approx \frac{\nabla_s^2}{n} \Rightarrow \mu = \sqrt{\frac{\nabla_s^2}{n}} \approx \sqrt{\frac{\nabla_0^2}{n}}$$

b. *pentru caracteristica alternativă*

Răspunsuri da / nu

Răspuns afirmativ DA, $x_1 = 1$

Răspuns negativ NU, $x_2 = 0$

Notăm cu : p = frecvenţa relativă a variantei x_1

q = frecvenţa relativă a variantei $x_2 \Rightarrow p + q = 1$

$$\bar{x}_a = \frac{\sum_{i=1}^{2} x_i n_i}{\sum_{i=1}^{2} n_i} = \frac{x_1 \cdot n_1 + x_2 n_2}{n} = x_1 f_1 + x_2 f_2 = 1 \cdot p + 0q = p$$

Scanned with OKEN Scanner

$$\nabla_0^2 = \frac{\sum_{i=1}^{2}\left(x_i - \bar{x}_a\right)^2 \cdot n_i}{\sum_{i=1}^{2} n_i} = \sum_{i=1}^{2}\left(x_i - \bar{x}_a\right)^2 f_i =$$

$$\left(x_1 - \bar{x}_a\right)^2 f_1 + \left(x_2 - \bar{x}_a\right)^2 \cdot f_2 = \left(1 - \bar{x}_a\right)^2 p + \left(- \bar{x}_a\right)^2 q =$$

$$= \left(1 - 2\bar{x}_a + \bar{x}_a^2\right) p + \bar{x}_a^2 q = p - 2\bar{x}_a\, p + x_a^2 (p + q) =$$

$$= p - 2p^2 + p^2 = p - p^2 = p(1 - p) = p \cdot q$$

$$\mu = \sqrt{\frac{\nabla_0^2}{n}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}; \quad p = \frac{M}{N}; \quad \mu = \sqrt{\frac{p(1-p)}{n}} \qquad (4.3)$$

M = răspunsuri afirmative din colectivitatea generală

N = volumul colectivității generale

Notăm m — răspunsuri afirmative din colectivitatea de selecție (eşantion)

n — volumul eşantionului

$$W = \frac{m}{n}$$

Rezultă $$\mu = \sqrt{\frac{w(1-w)}{n}} \qquad (4.4)$$

2. **Eroarea limită admisă** (eroarea maximă admisă)

Este produsul dintre coeficientul de siguranță „t“ şi eroarea medie de sondaj

a) pentru caracteristica nealternativă

$$\Delta x = t \cdot \mu = t \sqrt{\frac{\nabla_s^2}{n}} = t \sqrt{\frac{p(1-p)}{n}} \qquad (4.5)$$

b) pentru caracteristica alternativă

$$\Delta w = t \cdot \mu = t \cdot \sqrt{\frac{w(1-w)}{n}} \qquad (4.6)$$

Observație. Se acceptă în aplicații: t = 3

Scanned with OKEN Scanner

3. **Volumul eşantionului** (n)

Se calculează din *eroarea limită admisă* (4.5), (4.6) în funcție de coeficientul de siguranță „t", eroarea limită admisă (Δx, Δw) şi frecvența relativă a răspunsului afirmativ (p sau w) :

a) pentru caracteristică nealternativă

$$\Delta x^2 = t^2 \frac{\nabla_s^2}{n} \Rightarrow n = \frac{t^2 \nabla_s^2}{\Delta x^2} \qquad (4.7)$$

b) pentru caracteristica alternativă

$$\Delta w^2 = t^2 \frac{w(1-w)}{n} \Rightarrow n = \frac{t^2 w(1-w)}{\Delta w^2} \qquad (4.8)$$

Observație. Se acceptă uzual t = 3

Valorile coeficientului de siguranță

(Tabelul funcției Gauss — Laplace)

Probabilitate	Coef. de siguranță
0,95 (95%)	1,96
0,96060 (96,06%)	2,06
0,97074 (97,074%)	2,18
0,98019 (98,019%)	2,33
0,99068 (99,068%)	2,60
0,99505 (99,505%)	2,81
0,99549 (99,549%)	2,84
0,9965 (99,65%)	2,92
0,9973 (99,73%)	3,00
0,9980 (99,80%)	3,09
0,99825 (99,825%)	3,13
0,9988 (99,88%)	3,24
0,9990 (99,90%)	3,29
0,9994 (99,94%)	3,43
0,9996 (99,96%)	3,54
0,9999 (99,99%)	3,90

Sondajul simplu nerepetat

Se calculează aceiaşi indicatori (ca şi în cazul sondajului simplu repetat) folosindu-se aceleaşi relații cu menţiunea că ele se vor corecta cu coeficientul K :

Scanned with OKEN Scanner

$$K = \sqrt{\frac{N-n}{N-1}} \approx \sqrt{\frac{N-n}{N}} = \sqrt{1-\frac{n}{N}} \qquad (4.9)$$

Rezultă :

1. **Eroarea medie de sondaj** (μ)

a. *pentru caracteristica nealternativă*

$$\mu = \sqrt{\frac{\nabla_0^2}{n} \cdot \left(1-\frac{n}{N}\right)} \approx \sqrt{\frac{\nabla_s^2}{n} \cdot \left(1-\frac{n}{N}\right)} \qquad (4.10)$$

b. *pentru caracteristica alternativă*

$$\mu = \sqrt{\frac{p(1-p)}{n} \cdot \left(1-\frac{n}{N}\right)} \approx \sqrt{\frac{w(1-w)}{n} \cdot \left(1-\frac{n}{N}\right)} \qquad (4.11)$$

2. **Eroarea limită admisă**

a) *pentru caracteristica nealternativă*

$$\Delta x = t \cdot \mu = t\sqrt{\frac{\nabla_0^2}{n}\left(1-\frac{n}{N}\right)} \cong t \cdot \sqrt{\frac{\nabla_s^2}{n}\left(1-\frac{n}{N}\right)} \qquad (4.12)$$

b. *pentru caracteristica alternativă*

$$\Delta w = t \cdot \mu = t\sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} \approx t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} \qquad (4.13)$$

3. **Volumul eşantionului**

a) *pentru caracteristica nealternativă*

$$n = \frac{t^2\nabla_0^2}{\Delta x^2 + \frac{t^2\nabla_0^2}{N}} \approx \frac{t^2\nabla_s^2}{\Delta x^2 + \frac{t^2\nabla_s^2}{N}} \qquad (4.14)$$

b) *pentru caracteristica alternativă*

$$n = \frac{t^2 p(1-p)}{\Delta w^2 + \frac{t^2 p(1-p)}{N}} \approx \frac{t^2 w(1-w)}{\Delta w^2 + \frac{t^2 w(1-w)}{N}} \qquad (4.15)$$

Sondajul mecanic

Este apropiat de sondajul simplu şi generează erori mai mici. Formulele de calcul pentru indicatorii ce corespund erorilor medie de sondaj şi limita admisă sunt aceleaşi ca şi în cazul sondajului simplu.

Scanned with OKEN Scanner

Sondajul tipic

Sondajul tipic sau stratificat este utilizat în cazul studierii statistice a fenomenelor eterogene, când unele eşantioane formate pot fi nereprezentative motiv pentru care nu se poate aplica sondajul simplu sau cel mecanic

În cazul fenomenelor eterogene se procedează la separarea populaţiei în grupe calitativ omogene $N = N_1 + N_2 + \ldots + N_m$ (m = numărul grupelor) şi din fiecare grupă se formează subeşantioane de volume $n_1, n_2, \ldots, n_m$, a.î. eşantionul n are volumul $n_1 + n_2 + \ldots + n_m$

Se calculează :

— mediile de grupă $\bar{x}_{is}$, $i = \overline{1,m}$

— dispersiile de grupă ∇_{is}, $i = \overline{1,m}$

iar pe baza acestora se determină media întregului eşantion şi media colectivităţii generale. Influenţa *factorilor întâmplători* din cadrul grupelor este reflectată cu ajutorul dispersiilor parţiale.

- Pentru colectivitatea generală

$$\overline{\nabla_i^2} = \frac{\sum_{i=1}^{m} \nabla_i^2 \cdot N_i}{\sum_{i=1}^{m} N_i}, \qquad N = N_1 + \ldots + N_m = \sum_{i=1}^{m} N_i$$

- Pentru colectivitatea de sondaj

$$\overline{\nabla_{is}^2} = \frac{\sum_{i=1}^{m} \nabla_{is}^2 \cdot n_i}{\sum_{i=1}^{m} n_i}, \qquad n = n_1 + n_2 + \ldots + n_m = \sum_{i=1}^{m} n_i$$

Pentru calculul indicatorilor de sondaj tipic se înlocuieşte dispersia generală cu media dispersiilor parţiale ($\overline{\nabla_i^2}$).

Observaţie. Există mai multe tipuri de sondaj stratificat, dar forma cea mai uzuală este *sondajul tipic proporţional*, în care

$$\frac{n}{N} = \frac{n_i}{N_i}, \quad deci: \quad n_i = \frac{n}{N} \cdot N_i = K \cdot N_i, \quad K = \frac{n}{N}$$

În final : $n_i = K \cdot N_i \left(i = \overline{1,m}\right) \left(K = \frac{n}{N}\right)$

Scanned with OKEN Scanner

Calculul indicatorilor în sondajul tipic :

1. *Eroarea medie de sondaj*

a) *pentru caracteristica nealternativă*

$$\mu = \sqrt{\frac{\overline{\nabla_{is}^2}}{n}}$$

b) *pentru caracteristica alternativă*

$$\mu = \sqrt{\frac{w_i(1-w_i)}{n}} = \sqrt{\frac{\sum_{i=1}^{m} w_i(1-w_i)}{\sum_{i=1}^{m} n_i} \cdot \frac{1}{n}}$$

2. **Eroarea limită admisă**

a) *Pentru caracteristica nealternativă*

$$\Delta x = t\mu = t \cdot \sqrt{\frac{\overline{\nabla}_{is}^2}{n}}$$

b) *pentru caracteristica alternativă*

$$\Delta w = t\mu = t \cdot \sqrt{\frac{\sum_{i=1}^{m} w_i(1-w_i)n_i}{\sum_{i=1}^{m} n_i} \cdot \frac{1}{n}}$$

3. **Volumul eşantionului**

a) *pentru caracteristici nealternative*

$$n = n_1 + n_2 + \ldots + n_m = \sum_{i=1}^{m} n_i; \quad n_i = \frac{t^2 \overline{\nabla}_{is}^2}{\Delta x^2}$$

b) *pentru caracteristici alternative*

$$n = n_1 + n_2 + \ldots + n_m = \sum_{i=1}^{m} n_i; \quad n_i = \frac{t^2 \overline{w_i(1-w_i)}}{\Delta^2 w},$$

$$\text{unde } \overline{w_i(1-w_i)} = \frac{\sum_{i=1}^{m} w_i(1-w_i)n_i}{\sum_{i=1}^{m} n_i}$$

Scanned with OKEN Scanner

În cazul sondajului *tipic nerepetat*, formulele celor trei indicatori sunt aceleaşi ca şi în cazul sondajului tipic repetat, dar înmulțite cu coeficientul

$$K = \sqrt{\frac{N-n}{N-1}} \cong \sqrt{\frac{N-n}{N}} = \sqrt{1-\frac{n}{N}}$$

Exemplu : $$\Delta x = t\mu = t\sqrt{\frac{\overline{\nabla}^2_{is}}{n}\cdot\left(1-\frac{n}{N}\right)}$$

Sondajul de serii

Se utilizează în cazul colectivităților compuse din unități complexe, de sine stătătoare şi nu din unități simple. Unitățile complexe nu pot fi considerate grupe omogene, deoarece unitățile simple ce intră în componența lor au caracteristici proprii şi se deosebesc una de alta.

Se extrag prin sondaj simplu sau mecanic unitățile complexe în vederea formării unui eşantion.

Se înregistrează şi prelucrează datele unităților simple ce formează unitățile complexe în eşantion.

Se calculează :

— mediile de serie

— dispersia dintre serii (care va înlocui dispersia generală ∇_0^2)

$$\overline{\delta_x^2} = \frac{\sum_{i=1}^{r}\left(\bar{x}_i - \bar{\bar{x}}_s\right)^2}{r}$$ — dispersie simplă

unde r = număr de serii

$\bar{x}_i$ — media de serie

$\bar{\bar{x}}_i$ — media de eşantion

sau $$\overline{\delta}_x^2 = \frac{\sum_{i=1}^{m}\left(\bar{x}_i - \bar{\bar{x}}_i\right)^2 \cdot n_i}{\sum_{i=1}^{m} n_i}$$ — (dispersie ponderată)

Avem : $\nabla_0^2 = \overline{\nabla}^2 + \overline{\delta}_x^2$, *deci* $\overline{\delta}_x^2 \le \nabla_0^2$,

Scanned with OKEN Scanner

Erorile de reprezentativitate fiind în cazul sondajului de serii mai mici sau cel mult egale cu erorile ce se produc când se organizează un sondaj simplu.

- **Calculul indicatorilor la sondajul de serii**

1. *Eroarea medie de sondaj* (de reprezentativitate)

a) *Sondaj de serii repetat* $\mu = \sqrt{\frac{\overline{\delta_x^2}}{r}}$ (a)

b) *Sondaj de serii nerepetat* $\mu = \sqrt{\frac{\overline{\delta_x^2}}{r}\left(1 - \frac{r+1}{R-1}\right)}$ (b)

$\left(\mu = \sqrt{\frac{\overline{\delta_x^2}}{r} \cdot \frac{R-r}{R-1}}\right)$; R = numărul seriilor din colectivitatea generală,

r = numărul seriilor din colectivitatea de sondaj

2. **Eroarea limită admisă**

$\Delta x = t\mu$; $\Delta w = t\mu$; μ din relațiile (a) respectiv (b) la sondaj repetat sau nerepetat.

3. **Volumul eşantionului**

$r = \frac{t^2 \overline{\delta_x^2}}{\Delta^2 x}$ — pentru sondajul de serii repetat

$r = \frac{R\, t^2 \overline{\delta_x^2}}{(R-1)\Delta^2 x + t^2 \overline{\delta_x^2}}$ — pentru sondajul de serii nerepetat

Observație. Indicatorii prezentați se aplică şi în cazul sondajului de serii cu caracteristică alternativă, cu mențiunea că dispersia caracteristicii nealternative se înlocuieşte cu dispersia caracteristicii alternative.

Probleme

1. O populație originară este formată din 5 firme, fiecare având o cifră de afaceri trimestrială (în miliarde lei) de: 2,3,5,7,13. Se consideră toate eşantioanele care pot fi formate din două firme, prin extragere repetată şi apoi nerepetată din populația originară. Se cere:

Scanned with OKEN Scanner

a) media populației originare ;
b) abaterea medie pătratică a pop. originare ;
c) eșantioanele formate în cazul extragerii repetate și nerepetate ;
d) media distribuției mediei eșantioanelor în cazul extragerii repetate și nerepetate ;
e) abaterea medie pătratică a distribuției mediei eșantioanelor, adică *eroarea standard a mediei.*

Rezolvare

a) $\bar{x}_{CA} = \frac{2+3+5+7+13}{5} = 6\ mld.\ lei$

b) $\nabla^2_{CA} = \frac{\sum_{i=1}^{5}\left(x_i - \bar{x}_{CA}\right)^2}{5} =$

$$= \frac{(2-6)^2 + (3-6)^2 + (5-6)^2 + (7-6)^2 + (13-6)^2}{5}$$

$\nabla^2_{CA} = 15{,}2; \quad \Rightarrow \quad \nabla_{CA} \cong 3{,}9$ mld. lei

c) *extragere repetată*

(2,2)	(2,3)	(2,5)	(2,7)	(2,13)
(3,2)	(3,3)	(3,5)	(3,7)	(3,13)
(5,2)	(5,3)	(5,5)	(5,7)	(5,13)
(7,2)	(7,3)	(7,5)	(7,7)	(7,13)
(13,2)	(13,3)	(13,5)	(13,7)	(13,13)

— **extragere nerepetată** $C_5^2 = \frac{5!}{2!(5-2)!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{1\cdot 2\cdot 1\cdot 2\cdot 3} = 10$

(2,3)	(3,5)	(5,7)	(7,13)
(2,5)	(3,7)	(5,13)	
(2,7)	(3,13)		
(2,13)			

d) *mediile eșantioanelor și media distribuției mediei eșantioanelor*

Scanned with OKEN Scanner

— **extragere repetată**

mediile

2	2,5	3,5	4,5	7,5
2,5	3	4	5	8
3,5	4	5	6	9
4,5	5	6	7	10
7,5	8	9	10	13

media distribuției mediei eșantioanelor

$$x_{CAS,r} = \frac{suma\ tuturor\ mediilor}{25} = \frac{150}{25} = 6$$

ceea ce demonstrează faptul că $\bar{x}_{CAS,r} = \bar{x}_{CA}$

— **extragere nerepetată**

mediile

2,5	4	6	10
3,5	5	9	
4,5	8		
7,5			

media distribuției mediei eșantioanelor

$$x_{CAS,n} = \frac{suma\ tuturor\ mediilor}{10} = \frac{60}{10} = 6$$

care demonstrează că $\bar{x}_{CAS,n} = \bar{x}_{CA}$

e) *dispersia* ∇_s^2 *a distribuției mediei*

— **extragere repetată**

$$\nabla_{S,r}^2 = \frac{\sum_{i=1}^{25}\left(x_i - x_{CAS,r}\right)^2}{25} = 7,6$$

$$\nabla_{S,r} = \sqrt{7,6} \cong 2,75$$

Rezultă că pentru o populație finită

$$\nabla_{S,r}^2 = \frac{\nabla_0^2}{n}, \quad adică \quad 7,6 = \frac{\nabla_{CA}^2}{2}$$

$$7,6 = \frac{15,2}{2}$$

Scanned with OKEN Scanner

– extragere nerepetată

$$\nabla^2_{S,n} = \frac{\sum_{i=1}^{10}\left(x_i - x_{CAS,n}\right)^2}{10} = 5,7$$

$$\nabla_{S,n} = \sqrt{\nabla^2_{S,n}} = \sqrt{5,7} = 2,39$$

Se demonstrează faptul că

$$\nabla^2_{S,n} = \frac{\nabla^2_0}{n}\cdot\frac{N-n}{N-1}, \quad \text{adică } 5,7 = \underbrace{\frac{15,2}{2}\cdot\frac{5-2}{5-1}}_{5,7}$$

$N = 5; \qquad n = 2$

2. Se dau zece firme cu următoarele cifre de afaceri (în miliarde lei)

1; 1,5; 2,0; 1,8; 1,2; 1,6; 2,2; 2,5; 1,1; 2,1

Considerând că cele zece firme reprezintă un eşantion de 5% extras întâmplător şi repetat, *să se calculeze limitele* între care va fi cuprinsă cifra de afaceri medie în colectivitatea generală dată, dacă probabilitatea cu care se garantează rezultatul este 0,995.

Rezolvare.

n = 10 şi reprezintă 5% din colectivitatea generală. Rezultă : N = 200 firme.

Coeficientul de siguranţă : t = 3

Pentru sondaj simplu repetat :

$$\bar{x}_{CAS,r} = \frac{\sum_{i=1}^{10} x_i}{10} = 1,7 \text{ miliarde lei}$$

$$\nabla^2_{CAS,r} = \frac{\sum_{i=1}^{10}\left(x_i - x_{CAS,r}\right)^2}{10} = 0,233\cdot 10^{18} \quad \text{lei}^2$$

Eroarea medie de sondaj

$$\mu = \sqrt{\frac{\nabla^2_{CA_{s,n}}}{n}} = \sqrt{\frac{0,233}{10}} \text{ mld. lei} = 0,1526 \text{ miliarde lei}$$

Scanned with OKEN Scanner

- **Eroarea limită admisă**

$$\Delta x = t \cdot \mu = 3 \cdot 0{,}1526 \text{ miliarde} = 0{,}4579 \text{ miliarde}$$

Limitele între care va fi cuprinsă cifra de afaceri medie

$$\bar{x}_{CA_{s,n}} - \Delta x \le \bar{x}_{CA_{s,n}} \le \bar{x}_{CA_{s,n}} + \Delta x$$

$$(1{,}7 - 0{,}4579)\text{mld.lei} \le \bar{x}_{CA_{s,n}} \le (1{,}7 + 0{,}4579)\text{mld. lei}$$

$$1{,}2421 \text{ mld. lei} \le \bar{x}_{CA_{s,n}} \le 2{,}1579 \text{ mld. lei}$$

Observație : Dacă eşantionul a fost extras întâmplător nerepetat, atunci

$$\mu = \sqrt{\frac{\nabla^2_{CA_{s,n}}}{n}\left(1 - \frac{n}{N}\right)} = 0{,}1488 \text{ miliarde lei}$$

$N = 200; \quad n = 10$

$$\Delta x = t\mu = 3 \cdot 0{,}1488 = 0{,}4464 \text{ mld. lei}$$

$$1{,}2536 \le \bar{x}_{CA_{s,n}} \le 2{,}1463 \text{ mld. lei}$$

3. Cinci sute de firme au o cifră de afaceri medie de 1,02 mld. lei lunar şi o abatere pătratică medie de 1,00 mld. lei.

Determinați probabilitatea ca un eşantion de 100 de firme să aibă *o cifră de afaceri totală* :

a) cuprinsă între 90 mld. lei lunar şi 100 mld lei lunar
b) superioară lui 110 mld. lei lunar

Rezolvare

Pentru distribuția eşantionară $\bar{x}_{CAS} = \bar{x}_{CAO} = 1{,}02$ mld. lei lunar

şi $$\nabla_{CAS} = \frac{\nabla_0}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{1}{\sqrt{100}} \sqrt{\frac{500-100}{499}} = 0{,}09 \text{ mld. lei}$$

Considerăm o distribuție normală

a) Cifra de afaceri medie a celor 100 de forme este cuprinsă între $0{,}90 \frac{\text{mld. lei}}{\text{lună}}$ şi $\frac{1{,}00 \text{ mld. lei}}{\text{lună}}$ 0,90 şi 1,0 unități standard înseamnă:

$$z_1 = \frac{x - \bar{x}_S}{\nabla_S} = \frac{0{,}9 - 1{,}02}{0{,}09} = -1{,}33$$

$$z_2 = \frac{1 - 1{,}02}{0{,}09} = -0{,}2222$$

Scanned with OKEN Scanner

Probabilitatea căutată este egală cu : (aria de subgraficul distr. Normale cuprinsă între dreptele $z_1 = -1{,}33$ şi $z_2 = -0{,}2222$) sau (aria între $z_1 = -1{,}33$ şi 0) – (aria între $z_2 = -0{,}22$ şi 0) = 0,4082 –0,0871 = 0,3211

a) $p_1 = 0{,}3211$ sau $p_1 = 32{,}11\%$

b) Probabilitatea căutată este egală

cu $1 - (\text{aria de la } z = -\infty \text{ la } z = 0{,}89)$

Observaţie: $1{,}1 \frac{\text{mld. lei}}{\text{lună}}$ în *unităţi standard* reprezintă

$$z_3 = \frac{1{,}1 - 1{,}02}{909} = 0{,}89$$

Rezultă:

$$p_2 = 1 - (0{,}5 + 0{,}3133)\,0{,}1867$$

$$p_2 = 18{,}67\%$$

4. Un sondaj de opinie (simplu) pe un eşantion de 100 subiecţi ce reprezintă 4% din populaţia originară. Probabilitatea cu care se garantează rezultatele este de 0,9973.

Sondajul a constat în întrebarea:

„Consideraţi că legislaţia privind investiţiile favorizează pe investitorul străin comparativ cu cel autohton?"

Frecvenţa relativă a răspunsurilor afirmative a fost 55%.

Se cere:

a) eroarea medie de sondaj şi eroarea limită admisă, dacă sondajul este repetat, precum şi între ce limite se încadrează frecvenţa răspunsurilor afirmative la populaţia originară;

b) aceeaşi cerinţă ca şi la punctul (a) în cazul sondajului nerepetat.

Rezolvare:

Din tabelul de valori a funcţiei Gauss — Laplace rezultă că pentru o probabilitate de 99,73% coeficientul de siguranţă t = 3

a) **sondaj simplu repetat**

$w = 55\% = 0{,}55$

$n = 1000 \Leftrightarrow 4\%$ din N; $N = 25000$

indicatorii sondajului

- Eroarea medie de sondaj

$$\mu = \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0{,}55(1-0{,}53)}{1000}} = 0{,}01573$$

Scanned with OKEN Scanner

- Eroarea limită admisă

$$\Delta W = t \cdot \mu = 3 \cdot 0{,}01573 = 0{,}047196 \qquad (4{,}7196\%)$$

- Pentru colectivitatea generală (populația originară) frecvența răspunsurilor afirmative se încadrează între limitele:

$$(0{,}55 - 0{,}047196;\quad 0{,}55 + 0{,}047196) \text{ sau}$$
$$(0{,}503;\quad 0{,}597$$

prin urmare

$$50{,}3\% \le p \le 59{,}7\%$$

b. **Sondaj simplu nerepetat**

- Eroare medie de sondaj

$$\mu = \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}55(1-0{,}55)}{1000}\left(1 - \frac{1000}{25000}\right)} = 0{,}0154$$

$$\mu = 1{,}54\%$$

- Eroarea limită admisă

$$\Delta w = t \cdot \mu = 3 \cdot 0{,}0154 = 0{,}0462, \qquad \Delta w = 4{,}62\%$$

Pentru colectivitatea generală frecvența răspunsurilor se încadrează între limitele

$$(0{,}55 - 0{,}0462;\quad 0{,}55 + 0{,}0462) \text{ sau}$$
$$(0{,}504;\quad 0{,}596),$$

rezultă $50{,}4\% \le p \le 59{,}60\%$

5. Se efectuează un sondaj tipic (stratificat) pe un eşantion de 30 firme ce reprezintă 6% din populația originară, privind mărimea valorii adăugate brute (VAB). Probabilitatea cu care se garantează rezultatul este 99,73%.

Scanned with OKEN Scanner

În urma sondajului se obțin următoarele date :

VAB (mld. lei)	Nr. firme Frecvența obs.	VAB (mld. lei)	Nr. firme (Frecv. absolută)
1	2	13	1
2	2	15	1
2	2	15	1
3	2	16	1
5	2	17	1
6	3	19	1
7	3	20	2
8	1	21	2
9	1	22	1
10	1	25	1
11	1	–	–
12	1	–	–

Se cere :

limitele între care va fi cuprinsă valoarea adăugată brută în colectivitatea originală dacă : a) sondajul tipic este repetat

b) sondajul tipic este nerepetat

Rezolvare :

Deoarece VAB nu sunt omogene, se formează grupe de firme din eşantion, grupele deosebindu-se prin mărimea VAB de 5 mld. lei.

gr 1 VAB $\in (0,5]$

gr 2 VAB $\in (5,10]$

gr 3 VAB $\in (10,15]$

gr 4 VAB $\in (15,20]$

gr 5 VAB $\in (20,25]$

VAB mld. lei	x_i VAB medie	Frecventa abs. a firmelor
$(0,5]$	2,5	8
$(5,10]$	7,5	9
$(10,15]$	12,5	4
$(15,20]$	17,5	5
$(20,25]$	22,5	4
Total		30

Scanned with OKEN Scanner

Gr. 1

Media grupei 1

x_k VAB	n_k
1	2
2	2
3	2
4	0
5	2
	8

$$\overline{x_1} = \frac{\sum_{k=1}^{5} x_k n_k}{\sum_{k=1}^{5} n_k} = \frac{22}{8} = 2{,}75$$

Dispersia grupei 1

$$\nabla_1^2 = \frac{\sum_{k=1}^{5} \left(x_k - \overline{x_1}\right)^2 \cdot nk}{\sum_{k=1}^{5} n_k} = \frac{17{,}5}{8} = 2{,}1875$$

Gr. 2

x_k VAB	n_k
6	3
7	3
8	1
9	1
10	1
	9

$$\overline{x_2} = \frac{\sum_{k=1}^{5} x_k n_k}{\sum_{k=1}^{5} n_k} = \frac{166}{9} = 7{,}33$$

$$\nabla_2^2 = \frac{\sum_{k=1}^{5} \left(x_k - \overline{x_1}\right)^2 nk}{\sum_{k=1}^{5} n_k} = \frac{16{,}0001}{9} = 1{,}778$$

Gr. 3

x_k VAB	n_k
11	1
12	1
13	1
14	0
15	1
	4

$$\overline{x_3} = \frac{\sum_{k=1}^{5} x_k n_k}{\sum_{k=1}^{5} n_k} = \frac{51}{4} = 12{,}75$$

$$\nabla_3^2 = \frac{\sum_{k=1}^{5} \left(x_k - \overline{x_3}\right)^2 nk}{\sum_{k=1}^{5} n_k} = \frac{8{,}75}{4} = 2{,}188$$

Scanned with OKEN Scanner

Gr. 4

x_k VAB	n_k
16	1
17	1
18	0
19	1
20	2
	5

$$\overline{x_4} = \frac{\sum_{k=1}^{5} x_k n_k}{\sum_{k=1}^{5} n_k} = \frac{92}{5} = 18{,}4$$

$$\nabla_4^2 = \frac{\sum_{k=1}^{5} \left(x_k - \overline{x}_4\right)^2 \cdot nk}{\sum_{k=1}^{5} n_k} = \frac{13{,}2}{5} = 2{,}64$$

Gr. 5

x_k VAB	n_k
21	2
22	1
23	0
24	0
25	1
	4

$$\overline{x_5} = \frac{\sum_{k=1}^{5} x_k n_k}{\sum_{k=1}^{5} n_k} = \frac{89}{4} = 22{,}25$$

$$\nabla_5^2 = \frac{\sum_{k=1}^{5} \left(x_k - \overline{x}_5\right)^2 \cdot nk}{\sum_{k=1}^{5} n_k} = \frac{10{,}75}{4} = 2{,}687$$

Dispersiile de grupă

Grupa j	∇_j^2	n_j	$\overline{x_j}$
1	2,1875	8	2,75
2	1,778	9	7,33
3	2,188	4	12,75
4	2,64	5	18,4
5	2,687	4	22,25
		30	

Scanned with OKEN Scanner

Media dispersiilor de grupă

$$\overline{\nabla}^2 = \frac{\sum_{j=1}^{5} \nabla_j^2 \cdot n_j}{\sum_{j=1}^{5} n_j} = 2,207$$

Sondaj tipic repetat

- Eroarea medie de sondaj

$$\mu = \sqrt{\frac{\overline{\nabla}^2}{n}} = \sqrt{\frac{2,207}{30}} = 0,2712316$$

- Eroarea limită admisă

$$\Delta x = t \cdot \mu = 3 \cdot 0,2712316 = 0,8136948$$

$$t = 3$$

$$\Rightarrow \quad \bar{x} - \Delta x \le VAB \le \bar{x} + \Delta x$$

$$\bar{x} = \frac{\sum_{j=1}^{5} \bar{x}_j n_j}{\sum_{j=1}^{5} n_j} = \frac{2,75 \cdot 8 + 7,33 \cdot 9 + 12,75 \cdot 4 + 18,4 \cdot 5 + 22,25 \cdot 4}{30} =$$

$$= \frac{319,97}{30} = 10,665666$$

$$\bar{x} - \Delta x = 10,665666 - 0,8136948 = 9,851968 \text{ mld. lei}$$

$$\bar{x} + \Delta x = 11,479364 \text{ mld. lei}$$

Concluzie : $9,851968 \le \overline{VAB} \le 11,479364$

Sondaj tipic nerepetat

- **Eroarea medie de sondaj**

$$\mu = \sqrt{\frac{\overline{\nabla}^2}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{2,207}{30}\left(1 - \frac{30}{500}\right)} = 0,2629688$$

30 ..... 6%

N ..... 100% $\Rightarrow \quad N = \frac{3000}{6} = 500$

Scanned with OKEN Scanner

- **Eroarea limită admisă**

$$\Delta x = t \cdot \mu = 3 \cdot 0{,}2629688 = 0{,}788964$$

$$\bar{x} + \Delta x = 11{,}45563$$

$$\bar{x} + \Delta x = 9{,}877702$$

$$\Rightarrow \quad 9{,}877702 \text{ mld. lei} \leq \underset{\substack{\text{din} \\ \text{colectivitatea} \\ \text{generală}}}{\overline{VAB}} \leq 11{,}45563 \text{ mld. lei}$$

6. Se efectuează un sondaj simplu pe un eşantion de 45 de firme, care reprezintă 3% din colectivitatea generală, privind cifra de afaceri. Probabilitatea cu care se lucrează este de 0,9973. Se obţin următoarele date :

CA mld. lei	Nr. firme F.A	CA mld. lei	Nr. firme F.A
0,3	2	1,2	8
2,1	1	0,6	2
1,7	7	2,0	1
2,3	2	2,5	1
0,7	4	2,4	2
0,8	8	0,1	4
1,5	3	–	–

Total 45 firme

Se cer :

a) Gruparea datelor în cinci grupe egale
b) determinaţi valorile centrale ale fiecărei grupe (clase)
c) calculaţi media aritmetică şi dispersia generală
d) calculaţi coeficientul de asimetrie Pearson şi Fisher
e) calculaţi limitele între care va fi cuprinsă CA în colectivitatea generală dacă sondajul este repetat şi nerepetat

Scanned with OKEN Scanner

Rezolvare :

a), b)

CA mld. lei	Nr. firme F.A n_i	Valoarea centrală CA x_i
$(0 \div 0,5]$	6	0,25
$(0,5 \div 1,0]$	14	0,75
$(1,0 \div 1,5]$	11	1,25
$(1,5 \div 2,0]$	8	1,75
$(2,0 \div 2,5]$	6	2,25
Total	45	

$M_e = 1,136$ mld. lei

$M_0 = 0,8636$

c. $$\bar{x}_{a.} = \frac{\sum_{i=1}^{5} x_i \cdot n_i}{\sum_{i=1}^{5} n_i} = \frac{0,25 \cdot 6 + 0,75 \cdot 14 + 1,25 \cdot 11 + 1,75 \cdot 8 + 2,25 \cdot 6}{45} =$$

$$= \frac{53,25}{45} = 1,1833$$

$$\sigma_0^2 = \frac{\sum_{i=1}^{5} (x_i - \bar{x}_a)^2 \cdot n_i}{\sum_{i=1}^{5} n_i} = 0,38444; \quad \sigma_0 = \sqrt{\sigma_0^2} = 0,62$$

d. $$C_{asP} = \frac{\bar{x}_a - M_0}{\sigma_0} = 0,5156; \quad C_{asF} = \frac{3(\bar{x}_a - M_e)}{\sigma_0} = 0,22887$$

e. **Sondaj simplu repetat**

$\Delta x = 0,2772; \quad t = 3 \quad \bar{x}_a - \Delta x < \overline{CA} < \bar{x}_a + \Delta_x$

$0,906 \text{ mld. lei} < \overline{CA} < 1,46 \text{ mld. lei}$

Sondaj nerepetat

$\Delta x = 0,273; \quad 0,9103 \text{ mld. lei} \leq \overline{CA} \leq 1,4563 \text{ mld. lei}$

Scanned with OKEN Scanner

4.3. Aplicații

A_1. Formulați caracteristicile unei selecții.

Studiul populațiilor statistice este în general greu de realizat din cauza numărului mare de unități (subiecți) care le compun. Din această cauză, din populația statistică se extrage o parte, denumită selecție sau sondaj (uneori eşantion), din studiul căreia urmărim să determinăm anumite date ce privesc populația (originară).

Procesul de a obține date privind populația totală pornind de la informații incomplete poartă denumirea de interferență.

Se pune firesc întrebarea dacă dintr-o selecție (populația parțială) putem să tragem concluzii valabile la nivelul întregii populații statistice. Pentru a trage concluzii sigure va trebui ca selecția să fie reprezentativă, înțelegând prin aceasta că structura ei este aceeaşi sau aproape aceeaşi ca a populației originare (totale).

Se precizează că o selecție este reprezentativă dacă fiecare unitate din selecție se alege la întâmplare şi dacă fiecare unitate din selecție se alege la întâmplare şi dacă fiecare unitate din populația totală are aceeaşi probabilitate ca să fie aleasă în populația parțială (selecție).

Examinarea unui student la un examen este un exemplu de aplicarea teoriei selecției. Populația totală, în acest caz, este formată din cunoştințele studentului. Examinatorul selectează întrebări pentru verificarea cunoştințelor. Dacă selecția este reprezentativă examinarea este corectă.

Teoria selecției se bazează pe legea numerelor mari.

A_2. Selecție formată pe baza repartiției binominale

Considerăm o urnă în care se află bile albe şi negre într-o proporție necunoscută. Facem extracții succesive, cu reîntoarcerea bilei extrase înaintea următoarei extracții. Nu cunoaştem probabilitatea p ca dintr-o extracție să scoatem o bilă albă. Valoarea lui p se poate estima, efectuând un număr de n extracții cu reîntoarcerea bilei în urnă. Admitem că din cele n extracții am obținut ? bile albe, frecvența de apariție a bilei albe fiind $f = \frac{\alpha}{n}$.

Scanned with OKEN Scanner

Utilizăm teorema lui Bernoulli: Pentru o repartiție binominală avem: $\lim_{n\to\infty} P\left[\left|\frac{\alpha}{n}-p\right|<\varepsilon\right]=1, \varepsilon>0$, adică frecvența $f_n=\frac{\alpha}{n}$ pentru valori mari ale lui n este foarte apropiată de probabilitatea p, întrucât probabilitatea ca f_n să difere de p cu un număr ε oricât de mic, tinde către 1 când n tinde la infinit.

Se demonstrează că repartiția $\Psi=\frac{\frac{\alpha}{n}-p}{\sqrt{\frac{p(1-p)}{n}}}$ tinde, pentru $n\to\infty$, către repartiția normală.

Dacă $\Psi\cong N(0,1)$, avem (Anexa)

$P(-1{,}65 \le \Psi \le 1{,}65) = 0{,}90$
$P(-1{,}85 \le \Psi \le 1{,}85) = 0{,}94$
$P(-1{,}96 \le \Psi \le 1{,}96) = 0{,}95$
$P(-2{,}17 \le \Psi \le 2{,}17) = 0{,}97$
$P(-2{,}33 \le \Psi \le 2{,}33) = 0{,}98$
$P(-2{,}58 \le \Psi \le 2{,}58) = 0{,}99$
$P(-2{,}80 \le \Psi \le 2{,}80) = 0{,}995$
$P(-2{,}95 \le \Psi \le 2{,}95) = 0{,}997$
$P(-3{,}09 \le \Psi \le 3{,}09) = 0{,}998$
$P(-3{,}61 \le \Psi \le 3{,}61) = 0{,}9997$

Ținând seama de valoarea lui Ψ, formulele de mai sus pot fi scrise sub forma : $P\left(\frac{\alpha}{n}-\theta\sqrt{\frac{p(1-p)}{n}}\le p\le\frac{\alpha}{n}+\theta\sqrt{\frac{p(1-p)}{n}}\right)=\eta$

Unde θ = 1,65; 1,75; 1,85; 1,96; 2,17; 2,33; 2,58; 2,80; 2,95; 3,09; 3,61 corespunzător η = 0,90; 0,92; 0,94; 0,95; 0,97; 0.98; 0,99; 0,995; 0,997; 0,998; 0,9997.

Inegalitățile din paranteze se scriu sub forma $\left|p-\frac{\alpha}{n}\right|\le\theta\sqrt{\frac{p(1-p)}{n}}$ din care rezultă : $\left(1+\frac{\theta^2}{n}\right)p^2-2\left(\frac{\alpha}{n}+\frac{\theta^2}{2n}\right)p+\frac{\alpha^2}{n}\le 0$

Scanned with OKEN Scanner

Dacă $\frac{\alpha}{n}$ nu are valori foarte apropiate de 1 sau 0, se neglijează termenii în $\frac{\theta^2}{n}$ (care au ordinul mai mic decât ordinul celorlalți termeni) se obține intervalul

$$\frac{\alpha}{n}-\theta\sqrt{\frac{\frac{\alpha}{n}\left(1-\frac{\alpha}{n}\right)}{n}}\le p\le\frac{\alpha}{n}+\theta\sqrt{\frac{\frac{\alpha}{n}\left(1-\frac{\alpha}{n}\right)}{n}}$$ în care p se determină cu probabilitatea η.

Un asemenea interval se numeşte *interval de încredere*, valabil cu condiția $\frac{\theta^2}{n} \langle\langle\ 1$ din cauza aproximațiilor admise.

Pragul de exactitate este $(1-\eta)$, care se exprimă de regulă în % $[(1-\eta)\cdot 100\%]$.

Dacă nu este satisfăcută condiția $\frac{\theta^2}{n} \langle\langle 1$ atunci intervalul de încredere este : $\dfrac{\frac{\alpha}{n}+\frac{\theta^2}{n}-\theta\sqrt{\frac{\frac{\alpha}{n}\left(1-\frac{\alpha}{n}\right)}{n}}\left(1+\frac{\theta^2}{4\alpha\left(1-\frac{\alpha}{n}\right)}\right)}{1+\frac{\theta^2}{n}}$;

$$\frac{\frac{\alpha}{n}+\frac{\theta^2}{n}+\theta\sqrt{\frac{\frac{\alpha}{n}\left(1-\frac{\alpha}{n}\right)}{n}}\left(1+\frac{\theta^2}{4\alpha\left(1-\frac{\alpha}{n}\right)}\right)}{1+\frac{\theta^2}{n}}$$

Scanned with OKEN Scanner

Exemplu:

Un laborator de produse de ciocolată a executat 8000 bucăți de ciocolată, dintre care 1500 bucăți au greutatea mai mare decât limita superioară admisă.

Cu o probabilitate η.= 0,99 să se determine intervalul de variație a probabilității teoretice ca o bucată de ciocolată luată la întâmplare să depășească limita superioară de greutate admisă.

Rezolvare

$\alpha = 1500$; $n = 6000$; $\frac{\alpha}{n} = \frac{1500}{600} = 0,25$

$$P\left(\frac{\alpha}{n} - 2,58\sqrt{\frac{\frac{\alpha}{n}\left(1-\frac{\alpha}{n}\right)}{n}} \le p \le \frac{\alpha}{n} + 2,58\sqrt{\frac{\frac{\alpha}{n}\left(1-\frac{\alpha}{n}\right)}{n}} \right) = 0,99$$

$$0,25 - 2,58\sqrt{\frac{0,25(1-0,25)}{6000}} \le p \le 0,25 + 2,58\sqrt{\frac{0,25(1-0,25)}{6000}}$$

$0,25 - 0,0144 = p = 0,25 + 0,0144$

$0,2356 = p = 0,2644$

suntem siguri că probabilitatea teoretică căutată se găseşte sigur în intervalul (0,2356 – 0,2644).

A_3. **Selecție formată pe baza schemei hipergeometrice**

Pentru determinarea populației parțiale (a selecției) se aleg la întâmplare din populația originară (totală) unități, care pot intra în alcătuirea selecției decât o singură dată.

Schema probabilistică este reprezentată de o urnă cu N bile, din care N_1 bile albe şi N_2 bile negre ($N_1 + N_2 = N$). Din urnă se extrag n bile ($n < N$), fără reîntoarcerea bilei în urnă, după fiecare extragere. Notăm cu ? numărul de bile albe care se află printre cele n bile extrase.

Notând : $\frac{\alpha}{n} = p_0$, $p = \frac{N_1}{N_1 + N_2} = \frac{N_1}{N}$; $q = \frac{N_2}{N_1 + N_2} = \frac{N_2}{N}$ şi

ţinând seama de formula dispersiei pentru repartiţia hipergeometrică

Scanned with OKEN Scanner

$$D(X)=npq\left(1-\frac{n-1}{N-1}\right)\cong npq\left(1-\frac{n}{N}\right)$$

(aproximarea fiind valabilă pentru valori mari ale lui n şi N) avem :

$$P\left(p_0-\theta\sqrt{\frac{p_0(1-p_0)}{n}\left(1-\frac{n}{N}\right)}\le p\le p_0+\theta\sqrt{\frac{p_0(1-p_0)}{n}\left(1-\frac{n}{N}\right)}\right)=\eta,$$

unde $\eta=\frac{1}{2\Pi}\int_{-\theta}^{\theta}e^{-\frac{x^2}{2}}dx$.

A4. La o firmă au fost înregistrate pe parcursul unui an 2500 comenzi pentru livrarea de bunuri materiale.

O selecţie de 100 de comenzi analizate a dat următoarele rezultate:

Valoarea comenzii Milioane lei	Nr. comenzi
5	40
10	30
20	20
25	10

cu o probabilitate de 95 % să se estimeze

a) procentul şi numărul comenzilor cu valori mai mici sau egal cu 10 milioane lei

b) procentul şi numărul comenzilor cu valori mai mari sau egale cu 25 milioane lei.

Rezolvare

a) $p_0=\frac{70}{100}=0,7;\ q_0=1-0,7=0,3;\ n=100;\ N=2500$

$$P\left(p_0-1,96\sqrt{\frac{p_0q_0}{n}\left(1-\frac{n}{N}\right)}\le p_1\le p_0+1,96\sqrt{\frac{p_0q_0}{n}\left(1-\frac{n}{N}\right)}\right)=0,95$$

$$0,7-1,96\sqrt{\frac{0,7\cdot0,3}{100}\left(1-\frac{100}{2500}\right)}\le p\le 0,7+1,96\sqrt{\frac{0,7\cdot0,3}{100}\left(1-\frac{100}{2500}\right)}$$

Scanned with OKEN Scanner

$0,7 - 0,088 = p_1 = 0,7 + 0,088$

$0,6120 = p_1 = 0,7880$

Numărul comenzilor A, cu valori mai mici sau egale cu zece milioane lei rezultă din :

$$0,6120 \le p_1 = \frac{A_1}{N} \le 0,7880;\ 1530 \le A_1 \le 1970$$

b) $p_0 = \frac{10}{100} = 0,1;\ q_0 = 0,9;\ n = 100;\ N = 2500$

$$p_0 - 1,96\sqrt{\frac{p_0 q_0}{n}\left(1 - \frac{n}{N}\right)} \le p_2 \le p_0 + 1,96\sqrt{\frac{p_0 q_0}{1n}\left(1 - \frac{n}{N}\right)}$$

$$0,1 - 1,96\sqrt{\frac{0,1 \cdot 0,9}{100}\left(1 - \frac{100}{2500}\right)} \le p \le 0,1 + 1,96\sqrt{\frac{0,1 \cdot 0,9}{100}\left(1 - \frac{100}{2500}\right)}$$

$$0,1 - 0,057612 = p_2 = 0,1 + 0,057612$$

$$0,042388 = p_2 = 0,157612$$

$$0,042388 \le p_2 = \frac{A_2}{N} \le 0,157612;\ 106 \le A_2 \le 394$$

A5. În oraşul T. sunt 252370 locuinţe. În urma unei selecţii de 1500 locuinţe, s-a constatat că în 380 locuinţe familiile respective au mai mult de doi copii.Să se estimeze numărul de locuinţe în care familiile respective au mai mult de doi copii cu o probabilitate de 95 %.

Rezolvare :

Notăm : n – efectivul eşantionului; N – efectivul populaţiei originare (n = 1500; N = 252370) ;

$p_0 = \frac{\alpha}{n} = \frac{380}{1500}$, procentul (probabilitatea) ca o familie din eşantion să aibă mai mult de doi copii.

$$P(p_1 \le p \le p_0) = 95\%$$

$$P\left(\frac{\alpha}{n} - 1,96\sqrt{\frac{\frac{\alpha}{n}\left(1 - \frac{\alpha}{n}\right)}{n}} \le p \le \frac{\alpha}{n} + 1,96\sqrt{\frac{\frac{\alpha}{n}\left(1 - \frac{\alpha}{n}\right)}{n}}\right) = 95\ \%$$

Scanned with OKEN Scanner

$$0,2533-1,96\sqrt{\frac{0,2533(1-0,2533)}{1500}}\leq p\leq 0,2533+$$

$$+1,96\sqrt{\frac{0,2533(1-0,2533)}{1500}},$$

$$0,2533-0,022 = p = 0,2533 + 0,022.$$

Rezultă $0,2313 \leq p \leq 0,2753$; numărul de locuințe în care familiile respective au mai mult de doi copii (cu probabilitatea de 95 %):

$$0,2313 \leq \frac{N_1}{252370} \leq 0,2753;$$

$$58373 \leq N_1 \leq 69477$$

Cu o probabilitate de 95 % numărul locuințelor în care familiile au mai mult de doi copii „se află" în intervalul (58373–69477).

A_6. În două comune A și B avem următoarea situație :

În localitatea A la un număr de 1250 familii există 310 familii care posedă peste 10 hectare teren arabil, iar în localitatea B la 1900 familii, posedă peste 10 hectare teren arabil 470 familii. Se poate afirma că probabilitatea ca o familie să posede peste 10 hectare teren arabil este aceeași în fiecare comună ?

Rezolvare :

Notăm cu p probabilitatea ca o familie să posede peste 10 hectare teren arabil. În ambele comune sunt 3150 familii din care 780 familii posedă peste 10 hectare teren arabil. Se adoptă :

$$p_{01} = \frac{780}{3150} = 0,2476$$

Avem : $n_1 = 1250$, $\alpha_1 = 310$

$n_2 = 1900$, $\alpha_2 = 470$

$$p_{01} = \frac{\alpha_1}{n_1} = 0,248; \quad p_{01} = \frac{\alpha_{21}}{n_2} = 0,2474$$

Se consideră variabilele aleatoare p_{01}, p_{02} corespunzătoare la n_1 și n_2 extracții succesive dintr-o urnă cu probabilitatea p, cu reîntoarcerea bilei înaintea fiecărei extracții.

Calculăm dispersia $\sigma^2(p_{01} - p_{02})$

Scanned with OKEN Scanner

Avem :

$$\sigma^2(p_{01}-p_{02})=\sigma^2(p_{01})+\sigma^2(p_{02})=\frac{p(1-p)}{n_1}+\frac{p(1-p)}{n_2}=\frac{n_1+n_2}{n_1\cdot n_2}p(1-p)$$

Se demonstrează că variabila redusă

$$VR=\frac{\frac{\alpha_1}{n_1}-\frac{\alpha_2}{n_2}}{\sqrt{\frac{n_1+n}{n_1\cdot n_2}p(1-p)}}$$

urmează, la limită, legea normală N (0,1).

În situația din problemă avem :

$$VR=\frac{0{,}248-0{,}2474}{\sqrt{\frac{1250+1900}{1250\cdot 1900}\cdot 0{,}2476(1-0{,}2476)}}=\frac{0{,}0006}{0{,}015719}=0{,}03817$$

Cu o probabilitate egală cu 0,95 raportul (VR) trebuie să se găsească în intervalul (–1,96–1,96). În concluzie valorile p_{01} şi p_{02} nu diferă semnificativ.

Scanned with OKEN Scanner

4.4. Probleme. Teoria elementară a sondajului

Problema 4.1.

Un grup de şase firme, care formează o populaţie originară, au făcut investiţii de capital pe parcursul unui trimestru după cum urmează :

Numărul de ordine al firmei (i):	1	2	3	4	5	6
Capital investit (x_i): (miliarde lei)	2	3	4	6	7	8

Considerând eşantioane formate din două firme, prin extragere repetată şi extragere nerepetată, din populaţia originară, calculaţi :

a) media aritmetică a populaţiei originare ;

b) abaterea medie pătratică a populaţiei originare ;

c) media aritmetică a tuturor eşantioanelor şi media distribuţiei mediei eşantioanelor

d) abaterea medie pătratică a distribuţiei mediei eşantioanelor (abaterea standard a mediei eşantioanelor, adică eroarea standard a mediei).

Rezolvare :

$$\text{a. } \bar{x}_a = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{2+3+4+6+7+8}{6} = \frac{30}{6} = 5 \quad \text{(miliarde lei)}$$

$$\text{b. } \nabla_{CI} = \sqrt{\frac{1}{n}\sum_{i=1}^{6}\left(x_i - \bar{x}_a\right)^2} =$$

$$= \sqrt{\frac{1}{6}\left[(2-5)^2 + (3-5)^2 + (4-5)^2 + (6-5)^2 + (7-5)^2 + (8-5)^2\right]} =$$

$$= \sqrt{\frac{1}{6}(9+4+1+1+4+9)} = \sqrt{\frac{28}{6}} = 2{,}160247 \text{ miliarde lei.}$$

Cazul 1: Eşantioane din două firme prin extragere repetată

c1) pentru $i = \overline{1,6}$ se pot forma eşantioanele cu numerele :

(1,1); (1,2); (1,3); (1,4); (1,5); (1,6)
(2,1); (2,2); (2,3); (2,4); (2,5); (2,6)
(3,1); (3,2); (3,3); (3,4); (3,5); (3,6)
(4,1); (4,2); (4,3); (4,4); (4,5); (4,6)
(5,1); (5,2); (5,3); (5,4); (5,5); (5,6)
(6,1); (6,2); (6,3); (6,4); (6,5); (6,6)

Scanned with OKEN Scanner

Media fiecărui eşantion (miliarde lei) : $\bar{x}_i$

2	2,5	3	4	4,5	5
2,5	3	3,5	4,5	5	5,5
3	3,5	4	5	5,5	6
4	4,5	5	6	6,5	7
4,5	5	5,5	6,5	7	7,5
5	5,5	6	7	7,5	8

Media distribuției mediei eşantioanelor (în număr de 36)

$$\bar{x}_s = \frac{\text{suma tuturor mediilor}}{36} = \frac{180}{36} = 5$$

ceea ce demonstrează că :

media distribuției mediei eşantioanelor este egală cu media populației originare d1) $\nabla_s = \sqrt{\frac{1}{36}\sum_{i=1}^{36}\left(\bar{x}_i - \bar{x}_s\right)^2} = 1{,}527525$

Rezultatul obținut demonstrează egalitatea.

$\nabla_s^2 = \frac{\nabla_{CI}^2}{n_s}$ sau $\nabla_s = \frac{\nabla_{CI}}{\sqrt{n_s}}$, n_s – efectivul eşantionului;

$$\left(\nabla_s = 1{,}527520 = \frac{2{,}16024}{\sqrt{2}} = 1{,}527520\right)$$

Cazul 2 : Eşantioane formate din două firme prin extragere nerepetată.

C2) Numărul eşantioanelor este $C_6^2 = \frac{6!}{2!4!} = 15$, adică grupe formate din firme cu numerele :

(1,2);	(1,3);	(1,4);	(1,5);	(1,6);
(2,3);	(2,4);	(2,5);	(2,6);	
(3,4);	(3,5);	(3,6);		
(4,5);	(4,6);			
(5,6);				

Mediile eşantioanelor (miliarde lei) : $\bar{x}_i^*$

2,5;	3;	4;	4,5;	5
3,5;	4,5;	5;	5,5;	
5;	5,5;	6		
6,5;	7			
7,5;				

Scanned with OKEN Scanner

Media distribuției mediei eșantioanelor

$$\bar{x}_s^* = \frac{1}{15}\sum_{i=1}^{15} \bar{x}_i^* = 5{,}000,$$

ceea ce demonstrează că $\bar{x}_s^* = \bar{x}_a$, adică media distribuției mediei eșantioanelor (obținute prin extragere nerepetată) este egală cu media aritmetică a populației originare.

d2) Dispersia distribuției mediei eșantioanelor este

$$\nabla_s^{*2} = \frac{1}{15}\sum_{i=1}^{15}\left(\bar{x}_i^* - \bar{x}_s^*\right)^2 = 1{,}866667$$

iar abaterea mediei pătratică a distribuției mediei eșantioanelor

$$\nabla_s^* = \sqrt{\nabla_s^{*2}} = 1{,}366260$$

Rezultă egalitatea

$$\nabla_s^{*2} = \frac{\nabla_{CI}^2}{n} \cdot \frac{N-n}{N-1},$$

N=6 – efectivul populației originare,

n– efectivul eșantioanelor; $\Rightarrow \nabla_s^* = \frac{\nabla_{CI}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$; întradevăr :

$$\frac{\nabla_{CI}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{2{,}160247}{\sqrt{2}} \cdot \sqrt{\frac{6-2}{6-1}} = 1{,}366260 = \nabla_s^*$$

Problema 4.2.

Se dau zece firme care au următoarele cifre de afaceri (CA) :

Numărul firmei i:	1	2	3	4	5	6	7	8	9	10
Cifra de afaceri (miliarde lei) x_i	1,1	1,5	1,7	2,0	2,3	2,8	2,4	2,5	3,1	3,0

Se consideră că cele zece firme reprezintă un eșantion de 8% extras întâmplător repetat și nerepetat. Să se determine între ce limite este cuprinsă cifra de afaceri medie în colectivitatea generală dată (populația originară), dacă probabilitatea cu care se garantează rezultatul este 0,9987.

Scanned with OKEN Scanner

Rezolvare :

Avem n=10 (8% din populaţia originară N=125 firme) coeficientul de siguranţă corespunzător probabilităţii de 99,87% este t=3.

Cazul1 : Sondaj simplu repetat

— media aritmetică a eşantionului $\bar{x}_{as} = \frac{1}{10}\sum_{i=1}^{10} x_i = 2{,}24$ miliarde lei

— dispersia $\nabla_s^2 = \frac{1}{10}\sum_{i=1}^{10}\left(x_i - \bar{x}_{as}\right)^2 = 0{,}3924$

— eroarea medie de sondaj $\mu = \sqrt{\frac{\nabla_s^2}{n}} = 0{,}198091$

— eroarea limită admisă

$$\Delta_x = t \cdot \mu = 3 \cdot 0{,}198091 = 0{,}594273 \text{ miliarde lei}$$

— cifra de afaceri medie a colectivităţii originare $\overline{CA_0}$ este cuprinsă între $\left(\bar{x}_{as} - \Delta x, \bar{x}_{as} + \Delta x\right)$, adică : $\bar{x}_{as} - \Delta x \le CA_0 \le \bar{x}_{as} + \Delta x$,

$$2{,}24 - 0{,}594273 \le \overline{CA_0} \le 2{,}24 + 0{,}594273,$$

sau $1{,}645727 \le \overline{CA_0} \le 2{,}834273$ miliarde lei

Cazul 2 : Sondaj întâmplător nerepetat

— eroarea medie de sondaj

$$\mu = \sqrt{\frac{\nabla_s^2}{n}\left(1 - \frac{n}{N}\right)} = 0{,}190002$$

— eroarea limită admisă

$$\Delta x^* = t \cdot \mu = 3 \cdot 0{,}190002 = 0{,}570006;$$

rezultă

$$\bar{x}_{as} - \Delta x^* \le \overline{CA_0} \le \bar{x}_{as} + \Delta x^*;$$

$$2{,}24 - 0{,}570006 \le \overline{CA_0} \le 2{,}24 + 0{,}570006;$$

$$1{,}669994 \le \overline{CA_0} \le 2{,}810006 \quad \text{miliarde lei}$$

Problema 4.3.

O populaţie generală (originală) formată din şase firme, fiecare cu o valoare adăugată brută (VAB) trimestrială (miliarde lei) de:

Scanned with OKEN Scanner

Numărul firmei i:	1	2	3	4	5	6
VAB x_i (miliarde lei)	3	5	6	7	8	11

Considerând toate eşantioanele care pot fi formate din trei firme, prin extragere nerepetată, din populaţia originară, să se calculeze :

a) media aritmetică a populaţiei originare ;
b) abaterea medie pătratică a populaţiei originare ;
c) media distribuţiei mediei eşantioanelor ;
d) abaterea medie pătratică a distribuţiei eşantioanelor (eroarea standard a mediei).

Rezolvare :

a) media aritmetică a populaţiei originare

$$\bar{x}_a = \frac{1}{6}\sum_{i=1}^{6} x_i = 6{,}666667$$

b) abaterea medie pătratică a populaţiei originare

$$\nabla_0 = \sqrt{\frac{1}{6}\sum_{i=1}^{6}\left(x_i - \bar{x}_a\right)^2} = 2{,}494438$$

c) numărul eşantioanelor formate din trei firme, prin extragere nerepetată este $C_6^3 = \frac{6!}{3!3!} = 20$:

Scanned with OKEN Scanner

Numărul de ordine al firmelor din eşantion	Valoarea adăugată brută (miliarde lei)	Media aritmetică a eşantionului (miliarde lei) $\bar{x}_i$:
(1, 2, 3)	(3, 5, 6)	4,666667
(1, 2, 4)	(3, 5, 7)	5,000000
(1, 2, 5)	(3, 5, 8)	5,333334
(1, 2, 6)	(3, 5, 11)	6,333333
(1, 3, 4)	(3, 6, 7)	5, 333333
(1, 3, 5)	(3, 6, 8)	5, 666667
(1, 3, 6)	(3, 6, 11)	6, 666666
(1, 4, 5)	(3, 7, 8)	6,000000
(1, 4, 6)	(3, 7, 11)	7, 000000
(1, 5, 6)	(3, 8, 11)	7, 333333
(2, 3, 4)	(5, 6, 7)	6, 000000
(2, 3, 5)	(5, 6, 8)	6, 333333
(2, 3, 6)	(5, 6, 11)	7, 333333
(2, 4, 5)	(5, 7, 8)	6, 666667
(2, 4, 6)	(5, 7, 11)	7, 666666
(2, 5, 6)	(5, 8, 11)	8, 000000
(3, 4, 5)	(6, 7, 8)	7, 000000
(3, 4, 6)	(6, 7, 11)	8, 000000
(3, 5, 6)	(6, 8, 11)	8, 333333
(4, 5, 6)	(7, 8, 11)	8, 666667

— media distribuţiei mediei eşantioanelor $\bar{x}^* = \frac{1}{20}\sum_{i=1}^{20}\bar{x}_i = 6{,}666667$

($\bar{x}^* = \bar{x}_a$ – media aritmetică a populaţiei originare este egală cu media distribuţiei mediei eşantioanelor).

— abaterea medie pătratică a distribuţiei mediei eşantioanelor

$$\nabla_S^* = \sqrt{\frac{1}{20}\sum_{i=1}^{20}\left(\bar{x}_i - \bar{x}^*\right)^2} = 1{,}115600 \text{ miliarde lei}$$

Calculăm (n = efectivul eşantionului; N = efectivul populaţiei originare)

$$\frac{\nabla_0}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{2{,}494438}{\sqrt{3}} \cdot \sqrt{\frac{6-3}{6-1}} = 1{,}115600 = \nabla_S^*$$

Se verifică relaţia : $\nabla_s^* = \frac{\nabla_0}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

Scanned with OKEN Scanner

CAPITOLUL 5

ESTIMAȚII

5.1. Formularea problemei estimației

Ca urmare a *cercetării selective,* apar probleme când rezultatele sunt generalizate asupra întregii populații statistice.

- Este vorba despre *concluziile şi deciziile* care se iau asupra întregului atunci când se studiază doar o parte a acestuia. *Generalizările se referă la aproximarea parametrilor necunoscuți* care definesc legea de repartiție a caracteristicii aleatoare X studiate. (Estimarea parametrilor)
- Scopul cercetărilor este de a cunoaşte exact (cât mai exact) legea de repartiție.

Definiție. Dacă repartiția este exprimată printr-o funcție dată în care apar şi diverşi parametri necunoscuți, atunci avem o *repartiție specificată.*

Definiție. Dacă repartiția este exprimată printr-o funcție dată în care toți parametri sunt cunoscuți, atunci *avem o repartiție complet specificată.*

Fie $X_1, X_2, ..., X_n$ *o selecție independentă* de volum n *efectuată* asupra unei colectivități statistice a cărei caracteristică X ne interesează. Presupunem că repartiția lui X este dată de funcția $f(x,\theta)$, în care θ este un parametru scalar sau vector.

Observație.

— Dacă X este o variabilă aleatoare continuă (v.a.c.), atunci $f(x,\theta)$ este densitatea sa de repartiție.

— Dacă X este *variabila aleatoare discretă*, atunci $f(x,\theta)=P(X=x)$ reprezintă probabilitatea ca variabila aleatoare X să ia valoarea x.

Aplicație. Fie $X=$ *v.a.c.* cu

$$f(x,\theta)=\begin{cases}0 & x<0\\ \theta\cdot e^{-\theta\cdot x} & x\geq 0,\theta>0\end{cases}$$

Scanned with OKEN Scanner

atunci avem o repartiție continuă specificată. *Dacă aproximăm pe* θ, dându-i valoarea θ=2,5, atunci vorbim despre o repartiție complet specificată.

Aplicație.

Dacă X este v.a.d. (discretă) cu

$$f(x,\theta)=\begin{cases}0 & x\notin N\\ \dfrac{\theta^x}{x!}\cdot e^{-\theta} & x\in N\end{cases}$$

atunci *avem repartiție discretă specificată* (repartiția Poisson de parametru $\theta > 0$). Dacă estimăm pe θ dându-i valoarea θ_o=3,24, atunci avem repartiția discretă complet specificată şi vorbim despre o repartiție precisă dintr-o clasă anume.

Definiție. Fie $X_1, X_2, ..., X_n$ o selecție independentă de volum n asupra unei mulțimi a cărei caracteristică X are legea de repartiție $f(x,\theta)$. *În aceste condiții*:

- Orice funcție de variabilele aleatoare $X_1, X_2, ..., X_n$ pe care o notăm cu $T_n(X_1, ..., X_n)$ *se numeşte statistică.*
- Orice statistică $T_n(X_1, X_2, ..., X_n)$ se numeşte *funcție de estimație* sau estimator al parametrului θ al repartiției f(x,θ), dacă este utilizată în acest scop.
- Numărul $T_n(x_1, x_2, ..., x_n)$ sau valoarea estimatorului pentru sondajul efectiv $(x_1, x_2, ..., x_n)$ se numeşte estimație a lui θ, dacă a fost determinat cu acest scop.
- Orice procedeu de aproximare a unui parametru pe baza unui sondaj $(x_1, x_2, ..., x_n)$ cu ajutorul unui estimator se numeşte *estimare a parametrului considerat.*

Probleme :

— determinarea unei estimații (∃ o mulțime de estimatori a unui parametru);

— determinarea celei mai bune estimații a unui parametru, după anumite criterii de apreciere.

- Problema estimației unui parametru (unor parametri) *constă în determinarea unei valori numerice* sau a unei mulțimi de valori

Scanned with OKEN Scanner

care să acopere parametrul (sau parametrii consideraţi) aflat în studiu pe baza unei selecţii efectuate asupra populaţiei statistice, având caracteristica aleatoare X cu funcţia de repartiţie $f(x,\theta)$.

5.2. Tipuri de estimatori

- *Estimator absolut corect* al lui θ
- *Estimator corect* al lui θ
- Estimator nedeplasat al lui θ
- Estimator deplasat al lui θ

Tipuri de estimaţii

Estimaţia $T_n(x_1, x_2, ..., x_n)$ *este absolut corectă*, corectă, nedeplasată sau deplasată, după cum estimatorul $T_n(x_1, x_2, ..., x_n)$ este absolut corect, corect, nedeplasat sau deplasat.

- Statistica $T_n(x_1, x_2, ..., x_n)$ se numeşte :

1. *estimator absolut corect al lui* θ dacă

$$\begin{cases} M[T_n(X_1, X_2, ..., X_n)] = \theta \\ \lim\limits_{n\to\infty} D^2[T_n(X_1, X_2, ..., X_n)] = 0 \end{cases} \quad (5.1)$$

2. *estimator corect al lui* θ dacă

$$\begin{cases} M[T_n(X_1, X_2, ..., X_n)] = \theta + \Delta_n, \Delta_n \to 0 \\ \lim\limits_{n\to\infty} D^2[T_n(X_1, X_2, ..., X_n)] = 0 \end{cases} \quad (5.2)$$

Definiţie. Statistica $T_n(X_1, X_2, ..., X_n)$ se numeşte :

1. *estimator nedeplasat al lui* θ dacă

$$M[T_n(X_1, X_2, ..., X_n)] = \theta \quad (5.3)$$

2. *estimator deplasat al lui* θ dacă

$$M[T_n(X_1, X_2, ..., X_n)] = \theta + \Delta_n, \Delta_n \to 0 \text{ când } n\to\infty. \quad (5.4)$$

Observaţie. Diferenţa $\Delta_n = M(T_n) - \theta$ se numeşte *deplasare* şi putem avea *o estimare pozitiv deplasată* dacă $\Delta_n \rangle 0$ sau negativ deplasată dacă $\Delta_n \langle 0$.

Scanned with OKEN Scanner

Se mai utilizează : $\Delta_n \underline{\underline{def}}$ eroare sistematică şi nu trebuie confundată cu diferenţa $(T_n - \theta)$ care este o *variabilă aleatoare* şi care se numeşte *eroare de estimare*.

Propoziţie. Dacă $X_1, X_2, ..., X_n$ este o selecţie independentă de volum n asupra unei colectivităţi statistice având caracteristica x, cu valoarea medie $M(X) = m$, dispersia $D^2(X) = \sigma^2$ şi momentul iniţial de ordinul r $M(X^r) = Mr$, atunci :

- estimatorul $\overline{X}$ este absolut corect pentru parametrul m ;
- estimatorul S^2 este corect pentru parametrul σ^2 ;
- estimatorul $\tilde{S}^2$ este absolut corect pentru parametru σ^2 ;
- estimatorul $T_n(X_1, X_2, ..., X_n) = \frac{1}{n}\sum_{k=1}^{n} X_k$ este absolut corect pentru parametrul Mr; $\overline{X}$ = media de selecţie

Observaţie. Se remarcă faptul că: $\overline{X}$ este estimator nedeplasat al mediei teoretice m.

- $\tilde{S}^2$ este estimator nedeplasat al dispersiei teoretice σ^2.
- $\overline{Mr}$ (momentul iniţial de selecţie de ordinul r) este estimator nedeplasat al momentului iniţial teoretic de ordinul r (extinzând estimatorul $\overline{X}$).
- S^2 este estimator deplasat (negativ) pentru dispersia teoretică σ^2.

Definiţie. Se spune că statistica $T_n(X_1, X_2, ..., X_n)$ este *estimator consistent* al parametrului θ dacă $T_n(X_1, ..., X_n)$ converge în probabilitate către parametrul θ, adică:

$$(\forall)\varepsilon\rangle 0 \lim_{n\to\infty} P\left(\left|T_n(X_1, ..., X_n) - \theta\right|\langle\varepsilon\right) = 1 \qquad (5.5)$$

Propoziţie. Dacă $T_n(X_1, X_2, ..., X_n)$ *este un estimator corect* sau absolut corect al parametrului θ, atunci este şi consistent.

Propoziţie. Dacă $X_1, X_2, ..., X_n$ este o selecţie independentă de volum n asupra unei populaţii statistice având caracteristica X care admite momentele teoretice $M(X) = m$, $D^2(X) = \sigma^2$ şi $M(X^r) = Mr$, atunci:

Scanned with OKEN Scanner

- media de selecţie $\overline{X}$ este estimator consistent al mediei teoretice m;
- dispersia de selecţie S^2 şi dispersia de selecţie modificată $\tilde{S}^2$ sunt estimatori consistenţi pentru dispersia teoretică σ^2;
- momentul iniţial de selecţie de ordinul r, $\overline{Mr}$, este un estimator consistent pentru momentul iniţial teoretic de ordinul r.

Observaţie. Dacă $T_n(X_1, X_2, ..., X_n)$ este estimator consistent pentru θ, atunci $T_n(x_1, x_2, ..., x_n)$ este estimaţie consistentă pentru θ.

Observaţie. Fie X v.a.d. repartizată Poisson, cu

$$f(x,\theta) = \frac{\theta^x}{x!} \cdot e^{-\theta}, \qquad x \in N, \theta \rangle 0$$

Avem doi estimatori: $U_n(X_1, ..., X_n) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$V_n(X_1, ..., X_n) = \tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

cu $M(U_n) = \theta, \quad D^2(U_n) = 0 \quad (\text{când } n \to \infty)$

$M(V_n) = \theta, \quad D^2(V_n) = 0 \quad (\text{când } n \to \infty)$

Calculăm $D^2(V_n) - D^2(U_n) = \frac{1}{n}\left(3\theta^3 + \frac{2n}{n-1}\theta^2\right) \rangle 0$ ceea ce arată că $D^2(U_n) \langle D^2(V_n)$. Din cei doi estimatori corecţi ai parametrului θ, media de selecţie $\overline{X}$ este de minimă dispersie (criteriu de alegere).

Cantitate medie de informaţie pe observaţie

Definiţie. Fie X o variabilă aleatoare a cărei repartiţie este dată de funcţia $f(x,\theta)$continuă în raport cu θ şi derivabilă în raport cu θ de două ori. Expresia:

$$I(\theta) = M\left[\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2\right] = -M\left[\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}\right] \qquad (5.6)$$

se numeşte cantitate medie de informaţie pe observaţie în sens Fischer.

Observaţie. Dacă X este variabilă aleatoare continuă, atunci

$$I(\theta) = \int_{/R}\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2 \cdot f(x,\theta)dx = -\int_{/R}\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2} \cdot f(x,\theta)dx$$

Scanned with OKEN Scanner

iar dacă X este variabilă aleatoare discretă, atunci:

$$I(\theta)=\sum_i\left(\frac{\partial \ln f(x_i,\theta)}{\partial\theta}\right)^2\cdot f(x_i,\theta)=-\sum_i\frac{\partial^2\ln f(x_i,\theta)}{\partial\theta^2}\cdot f(x_i,\theta)$$

unde indicele de însumare poate parcurge o mulțime finită sau infinită și numărabilă.

Aplicație. Dacă X este repartizată binomial, atunci

$$f(x,\theta)=C_n^x\theta^x(1-\theta)^{n-x},\qquad x\in\{0,1,\dots,n\},\ 0\langle\theta\langle 1$$

$$\ln f(x,\theta)=\ln C_n^x+x\ln\theta+(n-x)\ln(1-\theta)$$

$$\frac{\partial\ln f(x,\theta)}{\partial\theta}=\frac{x}{\theta}-\frac{n-x}{1-\theta};\qquad \frac{\partial^2\ln f(x,\theta)}{\partial\theta^2}=\frac{x}{\theta^2}-\frac{n-x}{(1-\theta)^2}$$

Notăm

$$A(\theta)=\sum_i\left(\frac{\partial\ln f(x_i,\theta)}{\partial\theta}\right)^2\cdot f(x_i,\theta)_i;\ B(\theta)=-\sum_{i=1}^n\frac{\partial^2\ln f(x_i,\theta)}{\partial\theta^2}\cdot f(x_i,\theta)$$

Existența dispersiei minime pentru estimatori absolut corecți

PROPOZIȚIA RAO-CRAMER

Fie *X o variabilă aleatoare* a cărei repartiție este definită de funcția $f(x,\theta)$ continuă în raport cu θ și derivabilă de două ori în raport cu θ. Fie $X_1,X_2,\dots,X_n$ o selecție independentă de volum n asupra lui X și $T_n(X_1,X_2,\dots,X_n)$ un estimator absolut corect al lui θ. Atunci avem :

$$D^2[T_n(X_1,X_2,\dots,X_n)]\geq\frac{1}{nI(\theta)}\qquad\text{(Inegalitatea Rao-Cramer)}$$

cu egalitate dacă și numai dacă există o constantă C, care depinde de θ și de n, astfel încât aproape sigur, pentru un sondaj dat $(x_1,x_2,\dots,x_n)$ arbitrar,

$$\sum_{i=1}^n\frac{\partial\ln f(x,\theta)}{\partial\theta}=C[T_n(x_1,x_2,\dots,x_n)-\theta]$$

Observație. În cazul repartiției Poisson: $C=\frac{n}{\theta}$.

Definiție: Dacă $\left(T_n^k(X_1,X_2,\dots,X_n)\right)$, $1\leq k\leq s$, sunt s estimatori absolut corecți ai parametrului θ al repartiției $f(x,\theta)$, atunci vom spune

Scanned with OKEN Scanner

că $T_n^j(X_1, X_2, ..., X_n)$, $1 \le j \le s$ este *estimator eficient* al lui θ dacă:

$$D^2\left[T_n^j(X_1, X_2, ..., X_n)\right] = \min_{1 \le k \le s}\left\{D^2\left[T_n^k(X_1, X_2, ..., X_n)\right]\right\} \qquad (5.7)$$

Observație. Estimatorii eficienți sunt *estimatori de minimă dispersie*

Definiție. Dacă $T_n(X_1, X_2, ..., X_n)$ este *un estimator absolut corect* al parametrului θ al repartiției $f(x, \theta)$, atunci raportul

$$e(T_n) = \frac{\frac{1}{n \cdot I(\theta)}}{D^2(T_n)}, \qquad 0 \langle e(T_n) \le 1$$ se numeşte *coeficient de eficiență a estimatorului* T_n şi se spune că:

1. T_n este *estimator eficient* dacă $e(T_n) = 1$
2. T_n este *estimator asimptotic eficient* dacă $e(T_n) \to 1$ când $n \to \infty$.

Observație. $D^2(T_n)$ = măsoară eroarea de estimare a lui θ

Propoziție: Dacă U_n şi V_n sunt estimatori eficienți ai parametrului θ, atunci $U_n = V_n$ *aproape sigur* sau *aproape peste tot.*

5.3. **Estimare punctuală**

Metoda momentelor

Având ca suport teoretic faptul că *momentele de selecție inițiale de ordinul r sunt estimatori absolut corecți ai momentelor teoretice* inițiale de ordinul *r*, metoda momentelor (*propusă* de către englezul Karl Pearson (1857–1936), unul din fondatorii statisticii matematice) constă în :

- egalarea celor două tipuri de momente şi obținerea unui sistem de ecuații care să aibă atâtea ecuații câți parametri trebuie estimați.

Dacă X este o variabilă aleatoare având repartiția $f(x, \theta_1, \theta_2, ..., \theta_s)$, atunci se formează un sistem de forma :

$$M(X^r) = \overline{M}(X^r), \qquad 1 \le r \le s \qquad (5.8)$$

sau altfel spus : $M(X^r) = \frac{1}{n}\sum_{i=1}^{n} X_i^r$, $1 \le r \le s$ cu precizarea că ecuațiile trebuie să conțină toți parametrii de estimat.

Scanned with OKEN Scanner

- O soluție la sistemului (5.8) este acceptabilă dacă este de forma :

$$\hat{\theta}_r = \hat{\theta}_r(X_1, X_2, ..., X_n), \qquad 1 \le r \le s$$

ceea ce înseamnă că estimațiile sunt funcții de observații.

Observație. Metoda momentelor este aplicabilă numai pentru repartiții probabilistice care admit momente și care se pot calcula efectiv.

Observație. Pentru un același estimator putem avea mai multe estimații (valorile estimatorului pentru diferite sondaje concrete).

Aplicație. Dacă X este repartizată beta de parametri $\theta_1 = p$, $\theta_2 = q$, atunci :

$$f(x, \theta_1, \theta_2) = f(x, p, q) = \frac{1}{B(p,q)} x^{p-1} \cdot (1-x)^{q-1},$$

$$x \in [0,1], \qquad p \rangle 0, \quad q \rangle 0$$

Remarcând că : $M(X) = \frac{p}{p+q}$, $\qquad M(X^2) = \frac{p(p+1)}{(p+q)(p+q+1)}$

avem de soluționat sistemul de 2 ecuații cu 2 necunoscute :

$$M(X) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$M(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = S^2 + \overline{X}$$

obținând astfel estimatorii parametrilor p și q în funcție de momentele de selecție de ordinul întâi și doi.

$$T_1 = p(X_1, ..., X_n) = \frac{1}{S^2}\left[\overline{X}^2(1-\overline{X}) - S^2\overline{X}\right]$$

$$T_2 = q(X_1, ..., X_n) = \frac{1}{S^2}\left[\overline{X}(1-\overline{X}^2) - S^2(1-\overline{X})\right]$$

5.4. **Metoda verosimilității maxime**

Englezul R.A.Fischer (1890-1962), fondator al celei mai mari părți a statisticii matematice, a criticat metoda momentelor și *a propus* pentru același scop *metoda verosimilității maxime.*

- Fie *X o variabilă aleatoare continuă* cu densitatea de repartiție $f(x,\theta)$, $x \in R, \theta \in I \subseteq R$, derivabilă de două ori în raport cu θ. Fie, de

Scanned with OKEN Scanner

asemenea, $X_1, X_2, ..., X_n$ *o selecție independentă* de volum n asupra lui X. Variabilele aleatoare $X_1, ..., X_n$ sunt independente și identic repartizate cu X. O selecție particulară arbitrară $(x_1, x_2, ..., x_n)$ reprezintă o valoare a vectorului aleator $X_1, ..., X_n$ și este un punct curent $x = (x_1, x_2, ..., x_n)$ din spațiul de selecție R^n al variabilelor aleatoare $X_1, ..., X_n$.

- *Probabilitatea elementară* de realizare a vectorului $(x_1, x_2, ..., x_n)$ este:

$$P(x_1, x_2, ..., x_n, \theta)dx_1 dx_2, ..., dx_n = f(x_1, \theta) \cdot f(x_2, \theta)...f(x_n, \theta)dx_1...dx_n$$

Definiție. Funcția $P : R^n xI \to R$ dată de relația :

$$P(x;\theta) = P(x_1, x_2, ...x_n;\theta) = \prod_{i=1}^{n} f(x_i, \theta) = f(x_1, \theta)...f(x_n, \theta) \quad (5.9)$$

se numește funcție de verosimilitate.

Definiție. Fie $T_n(X_1, X_2, ..., X_n)$ o statistică având proprietatea că

$$[\max_{\theta}] P(x_1, x_2, ..., x_n;\theta) = P(x_1, x_2, ..., x_n; T_n(x_1, x_2, ..., x_n))$$

pentru fiecare selecție particulară $(x_1, x_2, ..., x_n)$. Atunci statistica $T_n(X_1, X_2, ..., X_n)$ *se numește estimator de maximă verosimilitate a parametrului* θ, iar valoarea sa particulară $T_n(x_1, x_2, ..., x_n)$ se numește estimație de maximă verosimilitate a parametrului θ.

Observație. În cazul unei variabile aleatoare discrete X, *funcția de verosimilitate reprezintă probabilitatea* de realizare a vectorului aleator $X_1, X_2, ..., X_n$ pentru a obține valoarea $(x_1, x_2, ..., x_n)$, adică:

$$P(X = x;\theta) = P[(X_1, X_2, ..., X_n) = (x_1, x_2, ..., x_n);\theta] =$$
$$= P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n;\theta)$$

și pentru că variabilele aleatoare sunt independente se ajunge la egalitatea

$$P(X_1, X_2, ..., X_n;\theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

Observație. Determinarea unui estimator de maximă verosimilitate pentru un parametru $\theta \in I \subseteq R$ se reduce la o problemă de extrem pentru funcții de o variabilă. Avem :

Scanned with OKEN Scanner

$$\frac{\partial P(x;\theta)}{\partial \theta} = P(x;\theta)\cdot\sum_{i=1}^{n}\frac{1}{f(x_i,\theta)}\cdot\frac{\partial f(x_i,\theta)}{\partial \theta} = P(x;\theta)\cdot\sum_{i=1}^{n}\frac{\partial \ln f(x_i,\theta)}{\partial \theta}$$

şi ca urmare căutarea punctelor staţionare se face prin rezolvarea ecuaţiei :

$$\frac{\partial P(x;\theta)}{\partial \theta} = 0 \quad \text{sau} \quad \sum_{i=1}^{n}\frac{\partial \ln f(x_i,\theta)}{\partial \theta} = 0 \qquad (5.10)$$

$$\frac{\partial^2 P(x;\theta)}{\partial \theta^2} = P(x;\theta)\left[\sum_{i=1}^{n}\frac{\partial \ln f(x_i,\theta)}{\partial \theta}\right]^2 + P(x;\theta)\sum_{i=1}^{n}\frac{\partial^2 \ln f(x_i,\theta)}{\partial \theta^2}$$

şi ca urmare, pentru un punct staţionar θ_o, semnul derivatei de ordinul doi este dat de expresia:

$$\frac{\partial^2 P(x_i;\theta)}{\partial \theta^2} = P(x_i;\theta\sum_{i=1}^{n}\frac{\partial^2 \ln f(x_i,\theta)}{\partial \theta^2}$$

în care avem în vedere faptul că $P(x;\theta_o)\rangle 0$.

În concluzie: Procedeul de determinare a *estimatorilor de maximă verosimilitate* în cazul unui singur parametru θ se prezintă astfel:

Pasul 1. Se determină funcţia de verosimilitate.

$$P(x;\theta) = P(x_1,x_2,...,x_n;\theta) = f(x_1;\theta)\cdot f(x_2;\theta)\cdot ...f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Pasul 2 : Se determină funcţia :

$$L(x,\theta) = \ln P(x;\theta) = \sum_{i=1}^{n}\ln f(x_i;\theta)$$

Pasul 3 : Se rezolvă ecuaţia de verosimilitate :

$$\frac{\partial L(x_i,\theta_0)}{\partial \theta} = 0 \text{ adică } \sum_{i=1}^{n}\frac{\partial \ln f(x_i,\theta)}{\partial \theta} = 0$$

ale cărei soluţii sunt punctele staţionare ale funcţiei de verosimilitate.

Pasul 4 : Fie $\theta_o = T_n(x_1,x_2,...,x_n)$ o soluţie a ecuaţiei de verosimi-litate, soluţia care trebuie să depindă de observaţiile $(x_1,x_2,...,x_n)$.

Dacă $\frac{\partial^2 L(x;\theta_o)}{\partial \theta^2}\langle 0$, atunci $\theta_o = T_n(x_1,x_2,...,x_n)$ este *o estimaţie de maximă verosimilitate pentru parametrul* θ, iar statistica

Scanned with OKEN Scanner

$T_n(X_1, X_2, ..., X_n)$ corespunzătoare este *un estimator* de maximă verosimilitate al lui θ.

Propoziție. Dacă $T_n(X_1, X_2, ..., X_n)$ este estimator eficient al parametrului θ, atunci acesta este şi estimator de maximă verosimilitate.

Aplicație. Dacă X este repartizată Poisson, adică

$$f(x,\theta) = \frac{\theta^x}{x!}e^{-\theta}, \quad x \in N, \theta \rangle 0$$

atunci pentru o selecție independentă de volum n, $(x_1, x_2, ..., x_n)$, avem funcția de verosimilitate:

$$P(x_1, x_2, ..., x_n; \theta) = \frac{\theta^{x_1}}{x_1!}e^{-\theta} \cdot \frac{\theta^{x_2}}{x_2!}e^{-\theta} ... \frac{\theta^{x_n}}{x_n!}e^{-\theta} =$$

$$= e^{-n\theta}\theta^{x_1+x_2+...+x_n} \cdot \frac{1}{x_1!x_2!...x_n!} = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} \cdot \frac{1}{\prod_{i=1}^{n} x_i!}$$

Să se aplice metoda verosimilității maxime:

(Paşii 2, 3, 4)

Pasul 2: $L(x;\theta) = \ln P(x;\theta) = -n\theta + \left(\sum_{i=1}^{n} x_i\right) \cdot \ln\theta - \ln \prod_{i=1}^{n} x_i$

Pasul 3: $\frac{\partial L(x,\theta)}{\partial\theta} = 0 \quad -n + \frac{1}{\theta}\left(\sum_{i=1}^{n} x_i\right) = 0 \quad \Rightarrow \theta_o = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \rangle 0$

Pasul 4: $\left.\frac{\partial^2 L(x,\theta)}{\partial\theta^2}\right|_{\theta=\theta_o} = -\frac{1}{\theta_0^2}\sum_{i=1}^{n} x_i = -\frac{n^2}{\sum_{i=1}^{n} x_i} \langle 0$

Concluzie: Media de selecție $\overline{X}$ este estimator de maximă verosimilitate, iar valoarea sa $\bar{x}$ pentru o selecție anume este estimație de maximă verosimilitate pentru parametrul θ al repartiției Poisson.

Aplicație : *Dacă X este repartizată normal* de parametrii $\theta_1 = m$ şi

$$\theta_2 = \sigma^2, \text{ adică } f(x,m,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right),$$

$$x \in R, m \in R, \sigma \rangle 0$$

Scanned with OKEN Scanner

atunci pentru o selecție independentă de volum n, $(x_1, x_2, ..., x_n)$, funcția de verosimilitate are expresia finală:

$$P(x_1, x_2, ..., x_n; \theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2\right] = P(x; m, \sigma)$$

Să se aplice metoda verosimilității maxime,
(Paşii 2, 3, 4)

Pasul 2 :

$$L(x, m, \sigma) = \ln P(x, m, \sigma) = -n \ln \sigma - n \ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

Pasul 3 :

$$\left.\begin{aligned} \frac{\partial L(x, m, \sigma)}{\partial m} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m)^2 = 0 \\ \frac{\partial L(x, m, \sigma)}{\partial \sigma} &= -\frac{m}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - m)^2 = 0 \end{aligned}\right\} \Rightarrow \begin{aligned} m_o &= \bar{x} \\ \sigma_o^2 &= s^2 \end{aligned}$$

Pasul 4 :

$$\frac{\partial^2 L}{\partial m^2} = -\frac{n}{\sigma^2}; \quad \frac{\partial^2 L}{\partial m \partial \sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{3}(x_i - m) = \frac{\partial^2 L}{\partial \sigma \partial m}$$

$$\frac{\partial^2 L}{\partial \sigma^2} = -\frac{n}{\sigma^2} - \frac{3}{\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

Obținem :

$$E(m_o, \sigma_o) = \frac{\partial^2 L(m_o, \sigma_o)}{\partial m^2} \cdot \frac{\partial^2 L(m_o, \sigma_o)}{2\sigma^2} - \left(\frac{\partial^2 L(m_o, \sigma_o)}{\partial m \partial \sigma}\right)^2 = \frac{2n^2}{\sigma_o^4} \rangle 0$$

Concluzie : Cum $\frac{\partial^2 L(m_o, \sigma_o)}{\partial m^2} \langle 0$, rezultă că statistica $(\bar{X}, S^2)$ este estimator de maximă verosimilitate, iar valoarea sa $(\bar{x}, s^2)$ este estimație de maximă verosimilitate pentru parametrul vectorial (m, σ^2) al repartiției normale $N(m, \sigma^2)$.

Propoziție : Fie X o variabilă aleatoare continuă cu densitatea de repartiție $f(x, \theta)$, $x \in R$, $\theta \in I \subset R$.

Scanned with OKEN Scanner

Fie, de asemenea, $(X_1, X_2, ..., X_n)$ o selecție independentă de volum n asupra lui X și $(x_1, x_2, ..., x_n)$ un rezultat al acesteia. În aceste condiții, dacă :

1. mulțimea I, a valorilor lui θ, este interval deschis al lui R

2. pentru orice $\theta \in I$ și orice $x \in R$, ($\exists$) derivatele parțiale $\frac{\partial \ln f}{\partial \theta}, \frac{\partial^2 \ln f}{\partial \theta^2}, \frac{\partial^3 \ln f}{\partial \theta^3}$

3. pentru orice $\theta \in I$

$$\left|\frac{\partial \ln f(x,\theta)}{\partial \theta}\right| < F_1(x), \left|\frac{\partial^2 \ln f(x,\theta)}{\partial \theta^2}\right| < F_2(x), \left|\frac{\partial^3 \ln f(x,\theta)}{\partial \theta^3}\right| < H(x,\theta)$$

funcțiile F_1 și F_2 fiind integrabile pe R, iar

$$\int_{/R} H(x,\theta) f(x,\theta) dx < K, \quad K = \text{independentă de } \theta.$$

4) $\int_{/R} \left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2 f(x,\theta) dx < \infty$ pentru orice $\theta \in I$ atunci

ecuația de verosimilitate are ca soluție un *estimator consistent* (și deci o estimație consistentă) al parametrului θ. Acest estimator este *asimptotic normal* și *asimptotic eficient.*

Scanned with OKEN Scanner

CAPITOLUL 6

IPOTEZE STATISTICE

6.1. Noțiuni generale

- Fie X o variabilă aleatoare unidimensională a cărei repartiție este dată de funcția f(x,θ) în care θ este un parametru real scalar sau vectorial.
- Orice presupunere privind repartiția necunoscută a unei variabile aleatoare X se numeşte ipoteză statistică.
- Ipotezele statistice pot fi formulate :

  — asupra formei (expresiei) funcției $f(x,\theta)$ sau

  — asupra valorilor parametrilor care apar în expresia funcției f(x,θ) considerată cunoscută.

  (În ultima variantă se spune că formulăm *ipoteze parametrice.*)
- presupunere de forma :

  $\theta = \theta_o$ sau $\theta \in A$ o notăm prin

  $H_o : \theta = \theta_o$ sau $H_o : \theta \in A$ (6.1)

Observație. O afirmație sau presupunere de forma $H_o : \theta = \theta_o$ sau $H_o : \theta \in A$ poate fi adevărată sau nu şi ca urmare, ipoteza H_o trebuie verificată.

Definiție. Notația $H_o : \theta = \theta_o$ sau $H_o : \theta \in A$ este denumită *ipoteză nulă, spre a o distinge de orice altă variantă* care se va numi *ipoteză alternativă.*

Observație. Dacă se pot formula mai multe presupuneri (6.1), atunci se spune că acestea sunt *ipoteze admisibile.*

Definiție. Presupunerea *punctuală* $H_o : \theta = \theta_o$ se numeşte ipoteză simplă. Presupunerea $H_o : \theta \in A$ se numeşte ipoteză compusă.

Scanned with OKEN Scanner

Modalitatea de verificare a ipotezelor. Regiune critică

Verificarea ipotezelor se face pe baza unei selecții de volum n asupra caracteristicii aleatoare X şi cu *ajutorul unei statistici* sau *funcții de selecție* aleasă corespunzător.

- Fie ipoteza nulă $H_0 : \theta = \theta_0$ şi ipoteza alternativă $H_1 : \theta = \theta_1$. Fie, de asemenea, *selecția de volum* n $(X_1, X_2, ..., X_n)$ şi $(x_1, x_2, ..., x_n)$ o valoare particulară a acesteia. Fie $T_n(X_1, X_2, ..., X_n)$ o statistică şi $\alpha \in (0,1)$ un număr suficient de mic ($\alpha = 0{,}01;\ 0{,}02;\ \alpha = 0{,}05$).
- Fie $V \subset R$ *mulțimea valorilor pe care le poate lua statistica* $T_n(X_1, X_2, ..., X_n)$ cu proprietatea că dacă ipoteza $H_o : \theta = \theta_o$ este adevărată, atunci:

$$P\left[T_n(X_1, X_2, ..., X_n) \in V \middle| H_o \text{adevărată}\right] = 1 - \alpha \qquad (6.2)$$

Observație. Mulțimii V cu proprietatea (6.2) îi corespunde o mulțime $W \subset R^n$, astfel încât

$$P\left[(x_1, x_2, ..., x_n) \in W \middle| H_o \text{adevărată}\right] = \alpha \qquad (6.3)$$

Definiție. Dacă există mulțimea W definită de (6.3), atunci $W \subset R^n$ se numeşte *regiune critică.*

Definiție. Numărul α se numeşte *nivel* sau *prag de semnificație.*

Observație. Presupunând că există regiunea critică W, iar numărul α este suficient de mic, atunci, pentru verificarea ipotezei $H_0 : \theta = \theta_0$, se procedează astfel:

1. dacă $(x_1, x_2, ..., x_n) \in W$ atunci *se respinge ipoteza* $H_0 : \theta = \theta_0$;
2. dacă $(x_1, x_2, ..., x_n) \notin W$ atunci *se acceptă ipoteza* $H_0 : \theta = \theta_0$.

Observație. Din modul în care se defineşte regiunea critică, se poate constata că aceasta *nu se determină în mod unic. Calea de obținere a regiunii critice reprezintă un procedeu de verificare a unei ipoteze statistice.*

- *Exemplu.* Fie $X \in N(m, \sigma)$ cu σ dat. Să presupunem că avem de verificat ipoteza $H_0 : m = m_0$. *Fie o selecție independentă* de volum

Scanned with OKEN Scanner

$n, (X_1, X_2, ..., X_n)$ efectuată asupra lui X şi $x = (x_1, x_2, ..., x_n)$ *o valoare de realizare* a acestei selecţii.

Considerăm statistica :

$$T_n(X_1, X_2, ..., X_n) = \frac{\sqrt{n}}{\sigma}(\overline{X} - m), \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(despre care ştim că este repartizată N(0,1). Fie $V = (-\infty, -a) \cup (a, \infty)$ şi ne propunem să determinăm pe $a \in R, a\rangle 0$, astfel încât :

$$P\left(T_n \in V \middle| H_o \text{adevărată}\right) = \alpha$$

- Dacă ipoteza $H_o : m = m_o$ este adevărată, atunci această egalitate are forma:

$$P\left(\frac{\sqrt{n}}{\sigma}\left|\overline{X} - m_o\right| \rangle a \middle| H_o \text{adevărată}\right) = \alpha$$ sau forma echivalentă:

$$P\left(-a \le \frac{\sqrt{n}}{\sigma}\left|\overline{X} - m_o\right| \le a \middle| H_o \text{adevărată}\right) = 1 - \alpha$$

ceea ce înseamnă ecuaţia în necunoscuta a :

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^{a} \exp\left(-\frac{t^2}{2}\right) dt = 1 - \alpha \qquad (6.4)$$

a cărei unică soluţie este $a = Z_{1-\frac{\alpha}{2}}$.

- Prin urmare, mulţimii V considerată mai sus îi corespunde regiunea critică următoare :

$$W = \left\{(x_1, x_2, ..., x_n) \in R^2 \middle| \frac{\sqrt{n}}{\sigma}\left|\overline{x} - m_o\right| \rangle Z_{1-\frac{\alpha}{2}}\right\}, \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (6.5)$$

Observaţie. În exemplul considerat mai sus, se remarcă apropierea dintre metoda de verificare a ipotezei $H_o : m = m_o$ şi metoda de construire a unui interval de încredere pentru parametrul *m*, dar şi deosebirea dintre ele.

Scanned with OKEN Scanner

6.2. Erori şi riscuri

Dacă $(x_1, x_2, ..., x_n) \in W$, atunci se va respinge ipoteza H_o, iar dacă $(x_1, x_2, ..., x_n) \notin W$, se va accepta H_o.

- Presupunem că $(x_1, x_2, ..., x_n) \in W$. Apar două posibilităţi :
  — să respingem H_o şi să acceptăm alternativa H_1 ;
  — să acceptăm ipoteza H_o.

(Aceleaşi posibilităţi se pot întâmpla şi în cazul $(x_1, x_2, ..., x_n) \notin W$, cu alte semnificaţii.)

Ca urmare, se pot întâmpla :

1. să se respingă H_o atunci când aceasta este adevărată, fapt care reprezintă *aşa-numita eroare de ordinul unu* şi care poate să apară cu probabilitatea α denumită şi *risc* de ordinul unu.

2. să se admită H_o atunci când aceasta este falsă, fapt care reprezintă *aşa-numita eroare de ordinul doi* şi care poate să apară cu probabilitatea β, denumită şi *risc de ordinul doi.*

În concluzie: În urma efectuării selecţiei de rezultat $(x_1, x_2, ..., x_n)$, ne putem afla într-una din următoarele situaţii:

$$P[(x_1, x_2, ..., x_n) \in W | \theta = \theta_o] = P[\text{se respinge } H_o | H_o \text{ adev.}] = \alpha$$

$$P[(x_1, x_2, ..., x_n) \notin W | \theta = \theta_o] = P[\text{se acceptã } H_o | H_o \text{ adev.}] = 1 - \alpha$$

$$P[(x_1, x_2, ..., x_n) \in W | \theta = \theta_1] = P[\text{se respinge } H_o | H_o \text{ falsã}] = 1 - \beta$$

$$P[(x_1, x_2, ..., x_n) \notin W | \theta = \theta_1] = P[\text{se acceptã } H_o | H_o \text{ falsã}] = \beta$$

Cu cât posibilităţile α şi β sunt mai mici, cu atât testul corespunzător este mai puternic.

6.3. Puterea unui test. Cel mai puternic test

Să presupunem că parametrul θ ia şi alte valori în afară de θ_o şi θ_1. *În acest caz, probabilitatea de respingere a ipotèzei* $H_o : \theta = \theta_o$ faţă de alternativa $H_1 : \theta = \theta_1$, *ca funcţie de* θ, *se numeşte funcţia de putere a testului* şi se notează prin $f_p(W, \theta)$.

Scanned with OKEN Scanner

Definiție. Valoarea funcției $f_p(W,\theta)$ pentru $\theta_1 \neq \theta_o$ dat se numește puterea testului față de alternativa θ_1.

Observație. Se poate scrie:

$$f_p(W,\theta) = P\left[(x_1, x_2, ..., x_n) \in W \middle| \theta\right] \text{ cu}$$

$$f_p(W,\theta_o) = \alpha; \qquad f_p(W,\theta_1) = 1-\beta.$$

Definiție. Expresia: $f_{co}(W,\theta) = 1 - f_p(W,\theta)$ reprezentând probabilitatea de acceptare a ipotezei H_o, se numește *funcție caracteristică operatorie* a testului, unde:

$$f_{co}(W,\theta_o) = 1-\alpha; \qquad f_{co}(W,\theta_1) = 1-\beta \qquad (6.6)$$

Observație. Un test este ideal dacă funcția caracteristică operatoare a testului este :

$$f_{co}(W,\theta_1) = \begin{cases} 1, & \text{dacă} \quad \theta = \theta_o \\ 0, & \text{dacă} \quad \theta = \theta_1 \end{cases} \qquad (6.7)$$

ceea ce presupune absența oricăror erori și prin urmare, absența oricăror riscuri.

În practica statistică, nu există un test ideal.

Se poate vorbi de teste mai bune sau mai puțin bune decât alte teste.

- Fie *nivelul de semnificație* sau riscul de ordinul unu α dat în așa fel, încât să aibă loc egalitatea :

$$P\left[(x_1, x_2, ..., x_n) \in W \middle| H_o : \theta = \theta_o\right] = \alpha \qquad (6.8)$$

Regula sau procedeul de alegere a regiunii critice W are în vedere relația (6.8) și dintre toate mulțimile $W \subset R^n$ care o satisfac, dacă există, este luată aceea pentru care numărul β, reprezentând riscul de ordinul doi și satisfăcând egalitatea:

$$P\left[(x_1, x_2, ..., x_n) \notin W \middle| H_1 : \theta = \theta_1\right] = \beta \qquad (6.9)$$

să ia valoarea cea mai mică posibilă.

Definiție. Dacă există un test astfel ca fixând sau păstrând probabilitatea erorii de ordinul întâi constantă, *alegem pe* W *ca bază de testare a ipotezei* $H_o : \theta = \theta_o$ *față de alternativa sa* $H_1 : \theta = \theta_1$, în așa fel încât să avem o probabilitate (un risc de ordinul doi) de valoare minimă pentru eroarea de ordinul doi, atunci *un astfel de test se numește cel mai puternic test.*

Scanned with OKEN Scanner

Aplicaţie. Fie $X \in N(m)$ cu σ dat şi dorim să verificăm ipoteza $H_o : m = m_o$. Fie $(X_1, X_2, ..., X_n)$ *o selecţie independentă* de volum n asupra lui X şi să notăm $x = (x_1, x_2, ..., x_n)$ *o valoare de realizare a selecţiei*.

Fie statistica :

$$T_n(X_1, X_2, ..., X_n; m) = \frac{\sqrt{n}}{\sigma}(\overline{X} - m); \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Deoarece statistica T_n urmează o repartiţie normală N(0,1), pentru un $\alpha \in (0,1)$ dat, putem determina două numere reale $a < b$ cu proprietatea:

$$\frac{1}{\sqrt{2\pi}}\int_a^b \exp\left(-\frac{t^2}{2}\right)dt = \phi(b) - \phi(a) = 1 - \alpha$$

unde ϕ = funcţia lui Laplace. Dacă ipoteza $H_o : m = m_o$ este adevărată atunci pentru α dat putem găsi numerele $a < b$, astfel încât:

$$P\left(a \le \frac{\sqrt{n}}{\sigma}(\overline{X} - m_o) \le b\right) = 1 - \alpha \qquad \text{sau}$$

$$P\left[(x_1, x_2, ..., x_n) \in R^n \middle| a \le \frac{\sqrt{n}}{\sigma}(\overline{x} - m_o) \le b\right] = 1 - \alpha$$

de unde deducem că:

$$P\left[(x_1, x_2, ..., x_n) \in R^n \middle| \frac{\sqrt{n}}{\sigma}(\overline{x} - m_o) \notin [a, b]\right] = \alpha$$

ceea ce înseamnă că regiunea critică este:

$$W = \left[(x_1, x_2, ..., x_n) \in R^n \middle| \frac{\sqrt{n}}{\sigma}(\overline{x} - m_o) \notin [a, b]\right] = W(a, \alpha)$$

înţelegând că pentru m_o şi σ daţi şi o selecţie efectuată, regiunea critică depinde de nivelul de semnificaţie α şi de numărul real a, numărul b putându-se determina în funcţie de a şi α din tabelele repartiţiei N(0,1).

Scanned with OKEN Scanner

- Să constatăm acum că :

$$P\left(a \le \frac{\sqrt{n}}{\sigma}(\overline{X} - m) \le b \right) =$$

$$= P\left(a - \frac{\sqrt{n}}{\sigma}(m_o - m) \le \frac{\sqrt{n}}{\sigma}(\overline{X} - m_o) \le b - \frac{\sqrt{n}}{\sigma}(m_o - m) \right)$$

sau altfel :

$$P\left[a \le \frac{\sqrt{n}}{\sigma}(\overline{x} - m) \le b \right] = \phi\left(b - \frac{\sqrt{n}}{\sigma}(m_o - m) \right) - \phi\left(a - \frac{\sqrt{n}}{\sigma}(m_o - m) \right)$$

de unde deducem că *dacă* ipoteza $H_o : m = m_o$ *este adevărată, atunci pentru o valoare arbitrară a parametrului* m şi pentru un nivel de semnificaţie α dat se determină a,b $(a < b)$, regiunea critică are forma:

$$W = \left\{ (x_1, x_2, \ldots, x_n) \in R^n \middle| \frac{\sqrt{n}}{\sigma}(\overline{x} - m_o) \notin \right.$$

$$\left. \notin \left[a - \frac{\sqrt{n}}{\sigma}(m_o - m), b - \frac{\sqrt{n}}{\sigma}(m_o - m) \right] \right\}$$

constatând că aceasta depinde de mai mulţi parametri.

Observaţie.

$$a \le \frac{\sqrt{n}}{\sigma}(\overline{x} - m) \le b \Leftrightarrow a \le \frac{\sqrt{n}}{\sigma}(\overline{X} - m_o + m_o - m) \le b \Leftrightarrow$$

$$a \le \frac{\sqrt{n}}{\sigma}\left[(-m + m_o) + (\overline{X} - m_o) \right] \le b - \frac{\sqrt{n}}{\sigma}(m_o - m)$$

$$a - \frac{\sqrt{n}}{\sigma}(m_o - m) \le \frac{\sqrt{n}}{\sigma}(X - m_o) \le b - \frac{\sqrt{n}}{\sigma}(m_o - m) \qquad (6.10)$$

Scanned with OKEN Scanner

6.4. Ipoteze asupra parametrilor repartiției normale

Ipoteza $H_o : m = m_o$ față de $H_1 : m = m_1 \neq m_o$ cu σ^2 cunoscută. Fie $X \in N(m, \sigma^2)$ cu σ cunoscută și $(X_1, X_2, ..., X_n)$ o selecție independentă de volum n efectuată asupra lui X.

Avem în vedere că : $\overline{X} \in N\left(m, \frac{\sigma^2}{n}\right)$ și $\frac{\sqrt{n}}{\sigma}(\overline{X} - m) \in N(0,1)$.

Presupunem că ipoteza $H_o : m = m_o$ este adevărată și ne propunem să determinăm o constantă $a > 0$, astfel încât :

$$P\left(\left|\overline{X} - m_o\right| > a \middle| m = m_o\right) = \alpha \qquad (6.11)$$

ceea ce ne conduce la regiunea critică :

$$W\left\{(x_1, x_2, ..., x_n) \in / R^n \left|\left|\overline{x} - m_o\right| > a\right.\right\} \qquad (6.12)$$

Din (6.11), rezultă că :

$$P\left(\left|\overline{X} - m_o\right| \leq a \middle| m = m_o\right) = 1 - \alpha \qquad (6.13)$$

și ținând cont de funcția de repartiție corespunzătoare lui $\Phi \in N(0,1)$, deducem :

$$F\left(\frac{a\sqrt{n}}{\sigma}\right) - F\left(-\frac{a\sqrt{n}}{\sigma}\right) = 1 - \alpha = 2F\left(\frac{a\sqrt{n}}{\sigma}\right) - 1 \qquad (6.14)$$

astfel că din tabelele repartiției normale N(0,1) obținem soluția :

$$\frac{a\sqrt{n}}{\sigma} = Z_{1-\frac{\alpha}{2}} \quad \text{sau} \quad a = \frac{\sqrt{n}}{\sigma} Z_{1-\frac{\alpha}{2}} \qquad (6.15)$$

Așadar, regiunea critică (6.12) are forma :

$$W\left\{(x_1, x_2, ..., x_n) \in R^n \left|\left|\overline{x} - m_o\right| > \frac{\sigma}{\sqrt{n}} Z_{1-\frac{\alpha}{2}}\right.\right\} \qquad (6.16)$$

ceea ce înseamnă că dacă eșantionul $(x_1, x_2, ..., x_n)$ aparține lui W vom respinge ipoteza $H_o : m = m_o$ la un nivel de semnificație α.

Scanned with OKEN Scanner

Observaţie. Notăm : $Z = \frac{\sqrt{n}}{\sigma}(\overline{X} - m)$ statistica (6.17)

utilizată mai sus şi cu :

$$Z_{calculat} = \frac{\sqrt{n}}{\sigma}\left|\overline{x} - m_o\right|;\ Z_{tabelat} = Z_{1-\frac{\alpha}{2}}$$

Concluzie

- Verificarea ipotezei $H_o : m = m_o$ faţă de alternativa

  $H_1 : m = m_1 \neq m_o$ se realizează astfel :

  1. Pentru o selecţie dată se calculează $\overline{x}$ şi nr. $Z_{calculat}$.
  2. Pentru un nivel de semnificaţie α dat se determină din tabele $Z_{1-\frac{\alpha}{2}}$.
  3. Se compară $Z_{calculat}$ cu $Z_{tabelat}$ şi decidem astfel :

  – dacă $Z_{cal} \langle Z_{tab}, \Rightarrow$ se acceptă H_o;

  – dacă $Z_{cal} \rangle Z_{tab}, \Rightarrow$ se respinge H_o;

  – dacă $Z_{cal} = Z_{tab}$, atunci se recomandă alte selecţii.

Observaţie. Notaţia Z (din 6.17: $Z = \frac{\sqrt{n}}{\sigma}(\overline{X} - m)$) utilizată pentru a verifica H_o, a dat testului denumirea de *testul Z.*

Exemplu. Fie n=25; $\overline{x}$=14,85; σ=1,5; α_1=0,05; α_2=0,025. Care dintre ipotezele $H_o : m = 15{,}15$; $H_o : m = 15{,}50$ este adevărată pentru fiecare nivel de semnificaţie considerat ?

Rezolvare

$$F(Z) = \frac{1}{2} + \phi(Z), \qquad \phi(Z) = \frac{1}{\sqrt{2\pi}}\int_o^Z e^{-\frac{x^2}{2}}dx$$

$$Z_{1-\frac{\alpha_1}{2}} = Z_{1-\frac{0{,}05}{2}} = Z_{0{,}975}$$

Scanned with OKEN Scanner

În tabelul lui $\phi(Z)$, la valorile funcției se adaugă 0,5.
Aria de sub curba normală standardizată de la 0 la z

z	0	1	2	3	4	5	6	7	8	9
.										
.										
.										
1,9	0,4713	0,4719	726	732	738	744	750	756	761	767

$0,975 = 0,5 + 0,4750 \rightarrow 1,96$

$$Z_{tabelat} = Z_{1-\frac{\alpha_1}{2}} = Z_{0,975} = 1,96$$

Dacă $H_o : m = 15,15$, atunci :

$$Z_{calculat} = \frac{\sqrt{n}}{\sigma}\left|\bar{x} - m_o\right| = \frac{\sqrt{25}}{1,5}\left|14,85 - 15,15\right| = 1$$

Concluzie: $Z_{calculat} \langle Z_{tabelat}$ și ipoteza $H_o : m = 15,15$ se acceptă cu un nivel de semnificație $\alpha = 0,05$.

Dacă $H_o : m = 15,15$, atunci :

$$Z_{calculat} = \frac{\left|14,85 - 15,50\right|5}{1,5} = \frac{13}{6} = 2,16... \rangle Z_{tabelat} = 1,96$$

și ipoteza $H_o : m = 15,15$ se respinge la un nivel de semnificație α=0,05.

Observație. Dacă $\alpha = 0,025$, atunci $Z_{tabelat} = Z_{1-\frac{\alpha}{2}} = Z_{0,9875} = 2,24$ și constatăm că la un nivel de semnificație $\alpha = 0,025$ ambele ipoteze sunt adevărate.

Observație. Din relația (6.16), constatăm că mulțimea complementară a regiunii critice W pe care o notăm cu $\overline{W}$, are forma:

$$\overline{W} = \left\{(x_1, x_2, ..., x_n \in R^n \left| \bar{x} - \frac{\sigma}{\sqrt{n}} Z_{1-\frac{\alpha}{2}} \le m_o \le \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{1-\frac{\alpha}{2}} \right.\right\} \quad (6.19)$$

Scanned with OKEN Scanner

- *În acest caz, verificarea ipotezei* $H_o : m = m_o$ este echivalentă cu verificarea faptului că valoarea presupusă a parametrului $m(m = m_o)$ se găseşte sau nu în intervalul de încredere corespunzător al cărui *coeficient de încredere* este $\delta = 1 - \alpha$.

- *Ipoteza* $H_o : m = m_o$ *faţă de* $H_1 : m = m_1 \neq m_o$ *cu* σ^2 *necunoscută*

Deoarece σ^2 nu este cunoscută, se consideră statistica:

$$T = T_n(X_1, X_2, ..., X_n; m) = (\overline{X} - m)\frac{\sqrt{n}}{S} \qquad (6.20)$$

unde : $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ şi $\overline{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

având în vedere că dacă $X \in N(m, \sigma)$, atunci statistica T_n urmează o repartiţie Student cu (n-1) grade de libertate.

Analog testului Z, presupunem ipoteza $H_o : m = m_o$ adevărată şi căutăm o constantă $a > 0$, astfel încât :

$P(|\overline{X} - m_o| \rangle a | m = m_o) = \alpha$, ceea ce ne conduce la regiunea critică :

$$W = \{(x_1, x_2, ..., x_n) \in / R^n | \overline{x} - m_o | \rangle a \}$$

Ca urmare, dacă ipoteza $H_o : m = m_o$ este adevărată, atunci acceptarea acesteia are loc cu probabilitatea :

$$P\left(-a\frac{\sqrt{n}}{\overline{S}} \leq (\overline{X} - m_o)\frac{\sqrt{n}}{\overline{S}} \leq a\frac{\sqrt{n}}{\overline{S}}\right) = 1 - \alpha \qquad (6.21)$$

Din tabelele repartiţiei Student cu *n* grade de libertate, deducem numere de forma $t_n; \frac{\alpha}{2}$ ca urmare a egalităţilor :

$$\int_{-\infty}^{-tn;\frac{\alpha}{2}} \varphi_n(x)dx = \int_{tn;\frac{\alpha}{2}}^{\infty} \varphi_n(x)dx = \frac{\alpha}{2} \qquad (6.22)$$

($\varphi_n(x)$ = densitatea repartiţiei Student) ; din (6.21) rezultă :

$$a\frac{\sqrt{n}}{s} = t_{n-1;\alpha/2} \qquad (6.23)$$

unde *s* este valoarea statisticii *S* pentru sondajul $(x_1, x_2, ..., x_n)$.

Scanned with OKEN Scanner

Regiunea critică va avea acum forma :

$$W = \left\{ (x_1, x_2, ..., x_n) \in / R^n \left| \left| \bar{x} - m_o \right| \rangle \frac{s}{\sqrt{n}} t_{n-1;\alpha/2} \right. \right\} \qquad (6.24)$$

sau altfel :

$$W = \left\{ (x_1, x_2, ..., x_n) \in R^n \left| \frac{\left| \bar{x} - m_o \right|}{s} \cdot \sqrt{n} \rangle t_{n-1;\alpha/2} \right. \right\} \qquad (6.25)$$

ceea ce înseamnă că, dacă eşantionul $(x_1, x_2, ..., x_n)$ aparţine lui W, atunci vom respinge ipoteza $H_o : m = m_o$ la un nivel de semnificaţie α (sau comiţând o eroare de ordinul unu al cărei risc măsurat este α sau în procente 100 α %).

Concluzie. Pentru a verifica ipoteza $H_o : m = m_o$, faţă de alternativa $H_1 : m = m_1 \neq m_o$, se procedează astfel:

1. pentru o selecţie dată, se calculează $\bar{x}$, $\bar{s}$ şi numerele :

$$t_{calculat} = \left| \bar{x} - m_o \right| \frac{\sqrt{n}}{s}; \quad t_{tabelat} = t_{n-1;\alpha/2}$$

(n, α=daţi)

2. se compară $t_{calculat}$ şi $t_{tabelat}$ şi se ia decizia :

- dacă $t_{calc.} \langle t_{tab.}$ atunci se acceptă H_o ;
- dacă $t_{calc.} \rangle t_{tab.} \Rightarrow$ se respinge H_o ;
- dacă $t_{calc.} = t_{tab.} \Rightarrow$ se recomandă alte selecţii.

Aplicaţie. Presupunem : .

$$n = 25; \bar{x} = 12,64; s^2 = 2,25; \alpha = 0,02; \alpha_2 = 0,002$$

Care dintre ipotezele : $H_o : m = 11,74$ sau $H_o : m = 13,36$ se acceptă?

Rezolvare :

$\alpha_1 = 0,02$

Din tabelul pentru repartiţia Student

$$t_{n-1;\alpha/2} \Rightarrow t_{24;0,01} = 2,797$$

Dacă $H_o : m = 11,74 \Rightarrow t_{calculat} = \left| \bar{x} - m_o \right| \frac{\sqrt{n}}{s} = \left| 12,64 - 11,74 \right| \frac{5}{1,5} = 3$

Scanned with OKEN Scanner

$t_{cal.} = 3 \rangle t_{tab.} = 2.797$

şi ipoteza $H_o : m = 11,74$ se respinge la un nivel de semnificaţie $\alpha_2 = 0,02$.

- Dacă $H_o : m = 13,36 \Rightarrow t_{cal.} = 2,4 \langle t_{tab.} = 2.797$ şi ipoteza $H_o : m = 14,47$ se acceptă la nivelul de semnificaţie 0,02.

Observaţie. Dacă $\alpha = 0,002$, atunci $t_{tab.} = t_{24;0,001} = 3,745$ şi se constată că ambele ipoteze sunt acceptate (la nivelul de semnificaţie 0,002).

Observaţie. Notaţia cu T a statisticii (6.20) utilizată pentru a verifica ipoteza H_o, considerată mai sus, a dat denumirea metodei de *testul t.*

Observaţie. Din relaţia (6.25), se constată că mulţimea complementară a regiunii critice W, notată $\overline{W}$, are forma :

$$\overline{W} = \left\{ (x_1, x_2, ..., x_n) \in R^n \middle| \bar{x} - \frac{s}{\sqrt{n}} t_{n-1;\alpha/2} \le m_o \le \bar{x} + \frac{s}{\sqrt{n}} t_{n-1,\alpha/2} \right\} \quad (6.26)$$

Verificarea ipotezei $H_o : m = m_o$ *faţă de alternativa sa* $H_1 : m = m_1 \neq m_o$ *este echivalentă cu a constata* dacă valoarea presupusă a lui m (adică dacă $m = m_o$) se găseşte sau nu în intervalul de încredere corespunzător.

Ipoteza $H_o : \sigma^2 = \sigma_o^2$ *faţă de alternativa* $H_1 : \sigma^2 = \sigma_1^2 \neq \sigma_o^2$

Se ştie că dacă $X \in N(m, \sigma)$, atunci statistica:

$$T_n^1(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad (6.27)$$

este repartizată hi-pătrat cu (n-1) grade de libertate, la fel ca şi statisticile :

$$T_n^2(X_1, X_2, ..., X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 ; T_n^3(X_1, X_2, ..., X_n) =$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^3$$

care *sunt estimatori* ai parametrului σ^2, presupunând că $(X_1, ..., X_n)$ este o selecţie independentă de volum n efectuată asupra lui X (primul este estimator corect şi al doilea estimator absolut corect).

Scanned with OKEN Scanner

Fie statistica :

$$T_n(X_1, X_2, ..., X_n; \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 = x^2 \qquad (6.28)$$

şi să căutăm două numere a > 0, b > 0, a < b, astfel încât dacă ipoteza $H_o : \sigma^2 = \sigma_o^2$ este adevărată, atunci să aibă loc relaţia :

$$P\left(x^2 \langle a \,|\sigma = \sigma_o\right) = P\left(x^2 \rangle b \,|\, \sigma = \sigma_o\right) = \frac{\alpha}{2} \qquad (6.29)$$

Având în vedere modul de utilizare a tabelelor repartiţiei x^2, deducem:

$$a = x_1^2 = x^2_{n-1;1-\frac{\alpha}{2}} \qquad \frac{1-\delta}{2} = \frac{\alpha}{2} \Rightarrow \alpha = 1-\delta; \delta = 1-\alpha$$

$$P\left(x^2 \rangle x_1^2\right) = \int_{x_1^2}^{\infty} \varphi_{n-1}(x)dx = 1 - \frac{\alpha}{2} \qquad (6.30)$$

$$b = x_2^2 = x^2_{n-1;\frac{\alpha}{2}}; \qquad P\left(x^2 \rangle x_2^2\right) = \int_{x_2^2}^{\infty} \varphi_{n-1}(x)dx = \frac{\alpha}{2} \qquad (6.31)$$

unde $\varphi_{n-1}(x)$ este densitatea de probabilitate a repartiţiei x^2 cu (n-1) grade de libertate, corespunzând lui $\overline{x}$ (6.28).

Din relaţiile (6.28) — (6.31), rezultă că regiunea critică *W* are forma:

$$W = \left\{(x_1, x_2, ..., x_n) \in / R^n \left| \frac{1}{\sigma_o^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \langle x_1^2 \text{ sau } \frac{1}{\sigma_o^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \rangle x_2^2 \right.\right\} \qquad (6.32)$$

iar complementara sa, notată $\overline{W}$ va fi:

$$\overline{W} = \left\{(x_1, x_2, ..., x_n) \in R^n \left| x_1^2 \le \frac{1}{\sigma_o^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \le x_2^2 \right.\right\} \qquad (6.33)$$

Ca urmare, dacă eşantionul $(x_1, x_2, ..., x_n)$ „cade în W", atunci se respinge ipoteza H_o, iar dacă eşantionul cade în complementara regiunii critice, atunci se acceptă ipoteza H_o la un nivel de semnificaţie α.

Concluzie. Pentru o selecţie dată $(x_1, x_2, ..., x_n)$, calculăm $X^2_{calculat} = \frac{1}{\sigma_o^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ şi reţinem că pentru a verifica ipoteza

Scanned with OKEN Scanner

$H_o : \sigma^2 = \sigma_o^2$ față de alternativa $H_1 : \sigma^2 = \sigma_1^2 \neq \sigma_o^2$ se procedează astfel:

1. Calculăm: $\bar{x}, \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$ și $x^2_{calculat}$.

2. Pentru n și α dați, se determină din tabele X_1^2 și X_2^2 și apoi se ia decizia de acceptare sau respingere a ipotezei $H_o : \sigma^2 = \sigma_o^2$.

3. Se compară $x^2_{calculat}$ cu X_1^2 și x_2^2:

dacă $x_1^2 \langle x^2_{calculat} \langle x_2^2$, atunci se acceptă H_o;

dacă $x^2_{calculat} \langle x_1^2$ sau $x^2_{calculat} \rangle x_2^2$, atunci se respinge H_o;

dacă $x^2_{calculat} = x_1^2$ (sau x_2^2), atunci se recomandă altă selecție.

Exemplu: Dacă $n = 25; \sum_{i=1}^{25} \left(x_i - \bar{x}\right)^2 = 1{,}1025$ și $\alpha = 0{,}02$, care dintre ipotezele următoare: a) $H_o : \sigma^2 = 0{,}16$, b) $H_o : \sigma^2 = 0{,}09$;

c) $H_o : \sigma^2 = 0{,}00225$ este adevărată?

Rezolvare:

Pentru $\alpha = 0{,}02$ *din tabelele repartiției* x^2, deducem:

$$x_1^2 = x^2_{n-1;1-\frac{\alpha}{2}} = x^2_{24;0{,}99} = 10{,}9(10{,}85)$$

$$x_2^2 = x^2_{n-1;1-\frac{\alpha}{2}} = x^2_{24;0{,}01} = 43{,}00(42{,}98)$$

dacă

$$H_o : 0{,}16 \Rightarrow x^2_{calculat} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = \frac{1}{0{,}16^2} \cdot 1{,}1025 = 6{,}89 \langle x_1^2 = 10{,}85$$

și H_o *se respinge*;

dacă $H_o : \sigma^2 = 0{,}09 \Rightarrow X^2_{calculat} = \frac{1}{0{,}09^2} \cdot 1{,}1025 = 12{,}25$, avem:

$x_1^2 \langle x^2_{calculat} \langle x_2^2$ și H_o *se acceptă*;

$(10{,}85 \langle 12{,}25 \langle 42{,}98)$

dacă $H_o : \sigma^2 = 0{,}0225 \Rightarrow x^2_{calculat} = 49 \rangle x_2^2$ și H_o *se respinge*.

Scanned with OKEN Scanner

Observație. Relația (6.33) se mai poate scrie :

$$\overline{W} = \left\{ (x_1, x_2, ..., x_n) \in / R^n \left| \frac{1}{x_2^2} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \le \sigma_o^2 \le \frac{1}{x_1^2} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 \right. \right\} \tag{6.34}$$

şi *deducem că verificarea ipotezei* $H_o : \sigma^2 = \sigma_o^2$ față de alternativa $H_1 : \sigma^2 = \sigma_1^2 \neq \sigma_o^2$ *poate fi redusă la a constata* dacă valoarea σ_o^2 a parametrului σ^2 se găseşte în intervalul de încredere corespunzător dispersiei σ^2.

6.5. **Teste de semnificație**

1. **Testul Z**

Inferența statistică asupra unui eşantion (studiul comparativ eşantion-populație)

1.1 Pentru studiul mediei scorurilor IQ ale elevilor de 15 ani s-a analizat un eşantion format din 180 de subiecți, media obținută de elevii testați fiind 102,8. Cunoscând că abaterea standard a scorurilor IQ din populația originară este 16, studiați posibilitatea ca media populației să fie mai mare decât 100 (pragul de semnificație α =0,02).

- Ipotezele statistice: (H_0): $m = 100$

  (H_1): $m > 100$, test Z unilateral

dreapta

- $Z_{calc} = \dfrac{\overline{m} - m}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{102{,}8 - 100}{\dfrac{16}{\sqrt{180}}} = 2{,}348$
- $Z_{cr}\Big|_{\substack{\alpha=0,02 \\ \text{unil. dr.}}} \equiv Z_{1-\alpha} = Z_{0,98} = 2{,}054$
- $Z_{calc} > Z_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptam } H_1 \end{cases}$, deci media populației este mai mare decât 100 (cu probabilitate de cel puțin 98 %).

Scanned with OKEN Scanner

1.2 Se cunoaşte că scorurile IQ au în general o distribuţie normală cu media 100 şi abaterea standard 15. S-a testat un eşantion de 150 elevi laureaţi la olimpiadele şcolare, obţinându-se media 108,5 pentru acest eşantion. Cu un grad de certitudine de 99%, se poate concluziona elevii respectivi provin dintr-o populaţie cu media semnificativ diferită de cea a populaţiei „normale" ?

- Ipotezele statistice: (H_0): $m = 100$

  (H_1): $m \neq 100$, test Z bilateral

- $Z_{calc} = \dfrac{\overline{m} - m}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{108,5 - 100}{\dfrac{15}{\sqrt{150}}} = 6,940$

- $Z_{cr}\Big|_{\text{bilat.}}^{\alpha=0,01} \equiv Z_{1-\frac{\alpha}{2}} = Z_{0,995} = 2,575$

- $Z_{calc} > Z_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptă } H_1 \end{cases}$, deci media populaţiei originare a elevilor în cauză este diferită de media populaţiei „normale".

1.3 Proba „flicker" constă în determinarea numărului de flashuri ale unei surse luminoase (într-o secundă) la care subiectul capătă senzaţia de lumină continuă. Pentru acest studiu s-a testat un eşantion format din 100 de subiecţi, obţinându-se media 39,2 şi abaterea standard 4,16 pentru eşantion. La pragul de semnificaţie $\alpha=0,005$ putem estima că media populaţiei este mai mică decât 40 ?

- Ipotezele statistice : (H_0): $m = 40$

  (H_1): $m < 40$, test Z unilateral stânga

- $Z_{calc} = \dfrac{\overline{m} - m}{\dfrac{S}{\sqrt{n}}} = \dfrac{39,2 - 40}{\dfrac{4,16}{\sqrt{100}}} = -1,923$

- $Z_{cr}\Big|_{\text{unilat. stg.}}^{\alpha=0,005} \equiv Z_{\alpha} = Z_{0,005} = -2,575$

- $Z_{calc} > Z_{cr} \Rightarrow \begin{cases} \text{nu respingem } H_0 \\ \text{nu acceptam } H_1 \end{cases}$, deci nu putem verifica ipoteza făcută.

Scanned with OKEN Scanner

1.4 Un eşantion aleator format din 250 femei a fost chestionat asupra necesităţii interzicerii intreruperilor de sarcina, 81 dintre acestea răspunzând afirmativ. Se poate concluziona că mai puţin de 40% din femei doresc interzicerea intreruperilor de sarcină? (α=0,05).

Frecvenţa caracteristicii în cadrul eşantionului: $\bar{f} = \frac{81}{250} = 0,324$

- Ipotezele statistice: (H_0): $f = 0,4$

  (H_1): $f < 0,4$, test Z unilateral stânga

- $Z_{calc} = \frac{\bar{f} - f}{\sqrt{\frac{f(1-f)}{N}}} = \frac{0,324 - 0,4}{\sqrt{\frac{0,4 \cdot (1-0,4)}{250}}} = -2,453$

- $Z_{cr}\big|_{\substack{\alpha=0,05 \\ \text{unilat. stg.}}} \equiv Z_\alpha = Z_{0,05} = -1,645$

- $Z_{calc} < Z_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{accept } H_1 \end{cases}$, deci ipoteza specifică este verificată cu o probabilitate de cel puţin 95 %.

Inferenţa statistică asupra a două eşantioane (independente sau asociate) — metode parametrice

1.5 Pentru a testa eficacitatea unei noi metode de studiu a literaturii s-au format două grupe de elevi: grupă de control „1“ (formată din 65 subiecţi) şi respectiv grupa experimentală „2“ (70 subiecţi). În urma experimentului, la testele finale grupa de control a obţinut media 6,95 cu abaterea standard 2,07, iar grupa experimentală media 7,75 cu abaterea standard 1,83. Noua metodă de studiu a condus la rezultate semnificativ diferite ? La ce nivel de încredere?

- Ipotezele statistice : (H_0) : $m_1 - m_2 = 0$

  (H_1) : $m_1 - m_2 \neq 0$, test Z bilateral

- $Z_{calc} = \frac{(\bar{m}_1 - \bar{m}_2) - (m_1 - m_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(6,95 - 7,75) - 0}{\sqrt{\frac{2,07^2}{65} + \frac{1,83^2}{70}}} = -2,372$

Scanned with OKEN Scanner

- $Z_{cr}\big|_{\text{bilat.}}^{\alpha=0,01} \equiv Z_{1-\frac{\alpha}{2}} = Z_{0,995} = 2,575$
- $-Z_{cr} < Z_{calc} < Z_{cr} \Rightarrow \begin{cases} \text{nu respingem } H_0 \\ \text{nu acceptam } H_1 \end{cases}$, deci la nivelul de încredere $\alpha=0,01$ nu putem preciza dacă rezultatele sunt semnificativ diferite.
- $Z_{cr}\big|_{\text{bilat.}}^{\alpha=0,02} \equiv Z_{1-\frac{\alpha}{2}} = Z_{0,99} = 2,326$
- $Z_{calc} < -Z_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptam } H_1 \end{cases}$, deci la nivelul de încredere $\alpha=0,02$ putem afirma că rezultatele sunt semnificativ diferite (cu probabilitate mai mare de 98%).

1.6 Aceeaşi problemă. Putem afirma că noua metodă îmbunătăţeşte rezultatele, în medie, cu cel puţin 0,5 puncte ? ($\alpha=0,05$).

- Ipotezele statistice : $(H_0): m_2 - m_1 = 0,5$

  $(H_1): m_2 - m_1 > 0,5$, test Z unilateral dreapta
- $Z_{calc} = \dfrac{(\overline{m}_2 - \overline{m}_1) - (m_2 - m_1)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \dfrac{(7,75 - 6,95) - 0,5}{\sqrt{\dfrac{2,07^2}{65} + \dfrac{1,83^2}{70}}} = 0,889$
- $Z_{cr}\big|_{\text{unil. dr.}}^{\alpha=0,05} \equiv Z_{1-\alpha} = Z_{0,95} = 1,645$
- deci nu putem preciza dacă rezultatele sunt îmbunătăţite cu cel putin 0,5 puncte.

1.7 Se doreşte studierea atitudinii studenţilor în privinţa efectelor prezenţei obligatorii la cursuri. S-au chestionat două eşantioane de studenţi — unul format din anul terminal, celălalt din primul an de studiu, cu urmatoarele rezultate:

Intrebare : „Consideraţi că introducerea prezenţei obligatorii la cursuri poate conduce la rezultate mai bune ? “

	Eşantion 1 (an terminal)	Eşantion 2 (primul an)
Volumul eşantionului	$N_1=110$	$N_2=165$
Număr de răspunsuri afirmative	61	65

Scanned with OKEN Scanner

Se poate afirma că atitudinea studenților față de această problemă se modifică semnificativ ? (α =0,01).

- Frecvența răspunsurilor afirmative în cadrul eșantioanelor :

$$\bar{f}_1 = \frac{61}{110} = 0{,}555 \qquad \bar{f}_1 = \frac{65}{165} = 0{,}394$$

- Ipotezele statistice : (H_0): $f_1 - f_2 = 0$

(H_1): $f_1 - f_2 \neq 0$, test Z bilateral

$$Z_{calc} = \frac{(\bar{f}_1 - \bar{f}_2) - (f_1 - f_2)}{\sqrt{\bar{f}^* \cdot (1 - \bar{f}^*)} \cdot \sqrt{\frac{N_1 + N_2}{N_1 \cdot N_2}}} =$$

$$) \frac{(0{,}555 - 0{,}394) - 0}{\sqrt{0{,}458 \cdot (1 - 0{,}458)} \cdot \sqrt{\frac{110 + 165}{110 \cdot 165}}} = 2{,}625$$

unde $\bar{f}^* = \frac{N_1 \cdot \bar{f}_1 + N_2 \cdot \bar{f}_2}{N_1 + N_2} = 0{,}458$

- $Z_{cr}\Big|_{\substack{\alpha=0{,}01 \\ bilat.}} \equiv Z_{1-\frac{\alpha}{2}} = Z_{0{,}995} = 2{,}575$

- $Z_{calc} > Z_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptam } H_1 \end{cases}$, deci atutudinea studenților față de frecvența la cursuri se modifică semnificativ în facultate.

Inferenta statistică asupra a două eșantioane asociate — teste neparametrice, eșantioane mari)

Proba Wilcoxon, N>20

1.8 Se efectuează un experiment de educație morală pe 25 de copii. Scorurile obținute (conform grilei de evaluare) înainte și după experiment (eșantioanele 1, respectiv 2) sunt prezentate în tabelul următor. Există diferențe semnificative între comportamentul copiilor înainte și după experiment ? (α = 0,05).

Scanned with OKEN Scanner

Eşantion 1	21	30	24	27	35	40	27	33	28	22	35	38	24	30	31	24	26	29	34	39	41	19	23	31	29
Eşantion 2	24	28	23	31	28	33	25	24	25	23	30	38	22	32	30	26	26	34	32	41	32	17	20	25	21
Diferența (1-2)	-3	+2	+1	-4	+7	+7	+2	+9	+3	-1	+5	0	+2	-2	+1	-2	0	-5	+2	-2	+9	+2	+3	+6	+8

- Ipotezele statistice :

(H_0) : diferențele dintre eşantioane sunt nesemnificative şi pot fi atribuite întâmplării (suma diferențelor pozitive este egală cu suma diferențelor negative, deci experimentul respectiv nu influențează semnificativ comportamentul)

(H_1) : diferențele dintre eşantioane sunt semnificative (experimentul influențează semnificativ comportamentul copiilor)

Se grupează diferențele (funcție de valoarea absolută a acestora) şi se elimină cazurile cu diferențe nule :

Diferența (val. abs.)	1	2	3	4	5	6	7	8	9
Frecvența	3	8	3	1	2	1	2	1	2
Rangul diferenței	22	16,5	11	9	7,5	6	4,5	3	1,5

Se atribuie ranguri diferențelor (în ordine descrescătoare, dacă există mai multe diferențe de aceeaşi valoare le atribuim media aritmetică a pozițiilor) şi se calculează suma rangurilor diferențelor minoritare. Se observă din primul tabel că diferențele minoritare sunt cele negative (7 diferențe negative față de 16 diferențe pozitive) :

Diferențe (minoritare)	-1	-2	-3	-4	-5
Frecvența (f_i)	1	3	1	1	1
Rangul diferenței (r_i)	22	16,5	11	9	7,5

$$R = \left(\sum R_i \cdot f_i\right)_{\text{dif. minoritare}} = 22 \cdot 1 + 16{,}5 \cdot 3 + \ldots = 99$$

Scanned with OKEN Scanner

Se calculează media şi abaterea standard a distribuţiei parametrului R:

$$m_R = \frac{N \cdot (N+1)}{4} = \frac{23 \cdot (23+1)}{4} = 138$$

$$\sigma_R = \sqrt{\frac{N \cdot (N+1) \cdot (2N+1)}{24}} = \sqrt{\frac{23 \cdot (23+1) \cdot (2 \cdot 23+1)}{24}} = 32{,}879$$

unde *N* este numarul de cazuri cu diferenţe diferite de 0 !

$$Z_{calc} = \frac{R - m_R}{\sigma_R} = \frac{99-138}{32{,}879} = -1{,}186$$

- Valoarea critică: $Z_{cr}\Big|_{\substack{\alpha=0,05 \\ bilat.}} \equiv Z_{1-\frac{\alpha}{2}} = Z_{0,975} = 1{,}96$
- $-Z_{cr} < Z_{calc} < Z_{cr} \Rightarrow \begin{cases} \text{nu respingem } H_0 \\ \text{nu acceptam } H_1 \end{cases}$ deci nu putem răspunde la întrebare.

Inferenţa statistică asupra a două eşantioane independente — teste neparametrice, eşantioane mari)

Testul U (Mann — Whitney)

2. Testul x^2

2.1 Testul x^2 pentru independenţa a două variabile (nivel nominal)

Testul x^2 pentru concordanţa dintre distribuţiile de frecvenţe teoretice şi cele experimentale

2.2 Pentru elaborarea strategiilor de creare a noi locuri de muncă în zona X este necesară cunoaşterea structurii populaţiei de şomeri din punct de vedere al calificării profesionale. Se presupune că această structură este următoarea:

Scanned with OKEN Scanner

Profesie (categorie)	Procent (teoretic)
1 – muncitori necalificaţi	20 %
2 – muncitori calificaţi	40 %
3 – agricultori	15 %
4 – studii superioare / domenii economice	15 %
5 – studii superioare / domenii tehnice	10 %

şi se doreşte verificarea acestei distribuţii de frecvenţe teoretice cu ajutorul unui eşantion de 800 de persoane, dintre care 125 se încadrează în categoria 1, 381 în categoria 2, 81 în categoria 3, 141 în categoria 4 şi respectiv, 72 în categoria 5.

Distribuţia teoretică este verificată experimental (α=0,01) ?

- Se organizează datele problemei, calculând frecvenţele teoretice (f_t) pentru fiecare categorie cu formula :

$$f_{t_i} = \frac{proc_i}{100} \cdot N = prop_i \cdot N$$

Categoria profesională	Frecvenţa observată (f_o) – număr de persoane –	Frecvenţa teoretică (f_t) – număr de persoane –	Diferenţe standard (D_{st_i})
1	125	$\frac{20}{100} \cdot 800 = 160$	–2,77
2	381	$\frac{40}{100} \cdot 800 = 320$	3,41
3	81	120	–3,56
4	141	120	1,92
5	72	80	–0,89
Total	800	800	

- Ipotezele statistice :

(H_0) : diferenţele dintre frecvenţele observate şi cele teoretice sunt nesemnificative şi pot fi atribuite întâmplării (distribuţia teoretică este corectă — există concordanţa cu datele experimentale)

Scanned with OKEN Scanner

(H_1) : diferenţele dintre frecevenţele observate şi cele teoretice sunt semnificative (distribuţia teoretică nu este corectă)

- $x^2_{calc} = \sum_{i=1}^{k} \frac{(f_{o_i} - f_{t_i})^2}{f_{t_i}} = \frac{(125-160)^2}{160} + \frac{(381-320)^2}{320} + ... = 36,43$
- $x^2_{cr}\big|_{\substack{\alpha=0,01 \\ (k-1)\ gr.\ lib.}} = x^2_{cr}\big|_{\substack{\alpha=0,01 \\ 4\ gr.\ lib.}} = 13,28$

$x^2_{calc} > x^2_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptam } H_1 \end{cases}$ deci distribuţia teoretică nu este corectă (nu este în concordanţă cu cea observată experimental).

Categoriile care au condus la acest rezultat (la care diferenţele între frecvenţa observată şi cea teoretică sunt mari) se identifică calculând diferenţele standard:

$$D_{st_i} = \frac{f_{o_i} - f_{t_i}}{\sqrt{f_{t_i}}}$$

Se observă că neconcordanţa dintre distribuţia teoretică şi cea observată experimental se datoreaza în special diferenţelor mari – mai mari decât 2 – la categoriile 1, 2, şi 3 (deci presupunerilor eronate cu privire la procentele acestor categorii).

Testul x^2 neparametric pentru compararea a „*k*“ eşantioane independente

Testul Kruskal – Wallis distribuţii asimetrice, cu amplitudinea variaţiei datelor mare

2.3 Interesează diferenţele dintre nivelul de autoapreciere al subiecţilor (grupaţi în 3 eşantioane : „subevaluare“, „echilibru“ şi „supraevaluare“) din punct de vedere al succesului profesional (evaluat pe o scală de la 1 — minim la 20 — maxim). Datele de evaluare a succesului profesional sunt prezentate în tabel. La nivelul de încredere α =0,05, se înregistrează diferenţe semnificative între eşantioane ?

Scanned with OKEN Scanner

Eşantion 1 (tendința de subevaluare)		Eşantion 2 (echilibru)		Eşantion 3 (tendința de supraevaluare)	
Scor	Rang	Scor	Rang	Scor	Rang
12	8,5	13	6,5	9	13,5
16	3	18	2	11	10,5
7	16	10	12	8	15
9	13,5	15	4	12	8,5
11	10,5	19	1	14	5
13	6,5				
$N_1=6$	$\sum R_1=58$	$N_2=5$	$\sum R_2=25,5$	$N_3=5$	$\sum R_3=52,5$

- Ipotezele statistice :

(H_0) : diferențele dintre eşantioane sunt nesemnificative şi pot fi atribuite întâmplării (grupele sunt omogene, provin din aceeaşi populație statistică sub aspectul succesului profesional)

(H_1) : diferențele dintre eşantioane sunt semnificative (cel puțin două eşantioane provin din populații statistice diferite în privința variabilei studiate – succesul profesional)

- $$\chi^2_{calc} = \left[\frac{12}{N\cdot(N+1)}\cdot\sum_{i=1}^{k}\frac{\left(\sum R_i\right)^2}{N_i}\right]-3\cdot(N+1)=$$

$$=\left[\frac{12}{16\cdot(16+1)}\cdot\left(\frac{58^2}{6}+\frac{25,5^2}{5}+\frac{52,5^2}{5}\right)\right]-3\cdot(16+1)=3,793$$

- $$x_{\substack{\alpha=0,05\\(k-1)\text{ gr. lib.}}} = x^2_{cr}\Big|_{\substack{\alpha=0,05\\2\text{ gr. lib.}}} = 5,99$$

- $x^2_{calc} < x^2_{cr} \Rightarrow \begin{cases}\text{nu respingem } H_0\\ \text{nu acceptam } H_1\end{cases}$ deci nu putem afirma că diferențele dintre eşantioane sunt semnificative.

Scanned with OKEN Scanner

Estimarea abaterii standard a populației originare

2.4 Dorim estimarea abaterii standard a coeficientului IQ al populației cu un grad de certitudine de cel puțin 95%, dispunând de rezultatele unui eșantion de 25 subiecți. Abaterea standard a eșantionului este 16,24. Care este intervalul de încredere (estimare) pentru abaterea standard a populației ?

- Nivelul de încredere : $\alpha = 1 - \frac{95}{100} = 0{,}05$
- Valorile critice : $x^2_{cr1} \equiv x^2_{\frac{\alpha}{2}}\Big|_{(N-1)\ gr.lib.} = x^2_{0,025}\Big|_{24\ gr.lib.} =$

$$x^2_{cr2} \equiv x^2_{1-\frac{\alpha}{2}}\Big|_{(N-1)\ gr.lib.} = x^2_{0,975}\Big|_{24\ gr.lib.} =$$

- Intervalul de estimare :

$$IE = \left(\frac{S\sqrt{N}}{x^2_{cr2}}; \frac{S\sqrt{N}}{x^2_{cr1}} \right) = \left(\frac{16{,}24\sqrt{25}}{\quad}; \frac{16.24\sqrt{.25N}}{\quad} \right)$$

3. **Testul T**

Inferența statistică asupra unui eșantion (studiul comparativ eșantion-populație)

3.1 Pentru studiul mediei timpului de reacție la stimuli acustici s-a testat un eșantion format din 60 subiecți, obținându-se media 106,5 ms și abaterea standard 13,05. Abaterea standard a timpului de reacție în populația originară este necunoscută. Media timpului de reacție al populației statistice din care provine eșantionul este mai mică decât 110 ? (pragul de semnificație α =0,01).

- Ipotezele statistice : $(H_0) : m = 110$

$(H_1) : m < 100$, test T unilateral stanga

Scanned with OKEN Scanner

- $T_{calc} = \dfrac{\overline{m} - m}{\dfrac{S}{\sqrt{n-1}}} = \dfrac{106,5-110}{\dfrac{13,05}{\sqrt{60-1}}} = -2,06$
- $T_{cr}\Big|_{\text{unil. stg.}}^{\alpha=0,01} \cong -2,39$
  $(n-1)$ gr.lib.
- $T_{calc} > T_{cr} \Rightarrow \begin{cases} \text{nu respingem } H_0 \\ \text{nu acceptam } H_1 \end{cases}$, deci nu putem aprecia dacă media populației este mai mică decât 110 (la acest prag de semnificație).

3.2 Studiem numărul de cadre prezentate într-o secundă la care subiecții capătă senzația de mișcare continuă a obiectelor de pe ecran, cu ajutorul unui eșantion de 40 persoane. Media determinată a eșantionului este 19,65 iar abaterea standard 2,03. Valoarea mediei populației este mai mică decât 24 ? (pragul de semnificație $\alpha = 0,005$). Care este intervalul de încredere pentru media populației ?

- Ipotezele statistice: $(H_0): m = 24$
  $(H_1): m < 24$, test T unilateral stânga
- $T_{calc} = \dfrac{\overline{m} - m}{\dfrac{S}{\sqrt{n-1}}} = \dfrac{19,65-24}{\dfrac{2,03}{\sqrt{40-1}}} = -13,382$
- $T_{cr}\Big|_{\text{unil. stg.}}^{\alpha=0,005} \cong -2,704$
  $(n-1)$ gr.lib.
- $T_{calc} < T_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptam } H_1 \end{cases}$, deci media populației este mai mică decât 24, cu certitudine de peste 99,5 %.
- Valorile critice :

$$T_{cr1} \equiv T_{\frac{\alpha}{2}}\Big|_{(N-1)\text{ gr. lib.}} = T_{0,0025}\Big|_{39\text{ gr. lib.}} \cong -3,127$$

$$T_{cr2} \equiv T_{1-\frac{\alpha}{2}}\Big|_{(N-1)\text{ gr. lib.}} = T_{0,9975}\Big|_{39\text{ gr. lib.}} \cong 3,127$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

• Intervalul de estimare :

$$IE = \left(\overline{m} + T_{cr1} \cdot \frac{S}{\sqrt{N-1}}; \overline{m} + T_{cr2} \cdot \frac{S}{\sqrt{N-1}} \right) =$$

$$= \left(19{,}65 - 3{,}127 \cdot \frac{2{,}03}{\sqrt{40-1}}; 19{,}65 + 3{,}127 \cdot \frac{2{,}03}{\sqrt{40-1}} \right) = (18{,}633; 20{,}666)$$

Inferența statistică asupra a două eşantioane independente

3.3 Se studiază gradul de susceptibilitate (vulnerabilitate) la manipularea prin mass media, comparativ pentru două grupe de vârstă : 20–30 ani şi respectiv 40–60 ani. Conform grilei de evaluare, scorurile sunt între 1 (vulnerabilitate nula) si 10 (maxima). Rezultatele studiului sunt prezentate în tabel :

	Eşantion 1 20–30 ani	Eşantion 2 40–60 ani
Volumul eşantionului	30	25
Media esantionului	7,93	6,15
Abaterea standard	1,43	1,65

Este adevărat că tinerii sunt, în medie, mai vulnerabili la manipulare cu cel puțin 1 punct (pe scala de evaluare) decât persoanele mature ? (α =0,05).

• Ipotezele statistice: (H_0): $m_1 - m_2 = 1$

(H_1): $m_1 - m_2 > 1$, test T unilateral dreapta

• $T_{calc} = \dfrac{(\overline{m}_1 - \overline{m}_2) - (m_1 - m_2)}{\sqrt{\dfrac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2 - 2}} \cdot \sqrt{\dfrac{N_1 + N_2}{N_1 \cdot N_2}}} = \dfrac{(7{,}93 - 6{,}15) - 1}{0{,}423} = 1{,}844$

• $T_{cr}\Big|_{\substack{\alpha=0,05 \\ \text{unil. dr.} \\ (N_1+N_2-2)\text{gr. lib.}}} = T_{cr}\Big|_{\substack{\alpha=0,05 \\ \text{unil. dr.} \\ 53 \text{ gr. lib.}}} = 1{,}675$

• $T_{calc} > T_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptam } H_1 \end{cases}$, deci ipoteza este verificată.

Inferența statistică asupra a două eşantioane asociate

3.4 În primul semestru studenții grupei 1 (25 subiecți) nu au fost obligați să participe la cursuri, media rezultatelor în prima sesiune fiind 7,30 cu abaterea standard 1,81. În semestrul II prezența a devenit obligatorie, media rezultatelor obtinute in a doua sesiune fiind 7,85 cu abaterea standard 1,33. Se poate afirma ca prezenta obligatorie conduce in general la rezultate mai bune ? (α =0,01).

• Ipotezele statistice: (H_0): $m_2 - m_1 = 0 \quad (m_1 = m_2)$

(H_1): $m_2 - m_1 > 0$, test T unilateral dreapta

• $T_{calc} = \dfrac{(\overline{m}_2 - \overline{m}_1) - (m_2 - m_1)}{\sqrt{\dfrac{S_1^2}{N} + \dfrac{S_2^2}{N}}} = \dfrac{(7{,}85 - 7{,}30) - 0}{\sqrt{\dfrac{1{,}81^2}{25} + \dfrac{1{,}33^2}{25}}} = 1{,}224$

• $T_{cr}\Big|_{\substack{\alpha=0,01 \\ \text{unil. dr.} \\ (N-1)\text{gr. lib.}}} = T_{cr}\Big|_{\substack{\alpha=0,01 \\ \text{unil. dr.} \\ 24\ \text{gr. lib.}}} \cong 2{,}492$

• $T_{calc} < T_{cr} \Rightarrow \begin{cases} \text{nu respingem } H_0 \\ \text{nu acceptam } H_1 \end{cases}$, deci, la acest nivel de încredere, nu putem concluziona că prezența obligatorie conduce la rezultate semnificativ îmbunătățite.

4. Test F

Analiza de variantă — unifactorială

4.1 Se recurge la un experiment pentru a determina efectul unui tranchilizant asupra timpului de reacție, bazat pe testarea a 4 grupuri de persoane (un grup de control şi 3 grupuri experimentale care primesc doze diferite). Doza de tranchilizant influențează semnificativ (la nivelul populației) timpul de reacție ? (α =0,05).

• Se organizează datele problemei şi se calculează suma valorilor, respectiv suma pătratelor valorilor din fiecare eşantion :

Scanned with OKEN Scanner

	Gr. experim. 1	Gr. experim. 2	Gr. experim. 3	Gr. control
Timpi de reacţie	69,72,81,75,70,76	76,81,84,80,82,79	80,88,91,79,85	68,70,67,73,71
Volumul eşantionului	$N_1=6$	$N_2=6$	$N_3=5$	$N_4=5$
Suma valorilor	$T_1=443$	$T_2=482$	$T_3=423$	$T_4=349$
Suma pătratelor	$\sum x_1^2=32807$	$\sum x_2^2=38759$	$\sum x_3^2=35891$	$\sum x_4^2=24383$
Media eşantionului	$\overline{X}_1=73,83$	$\overline{X}_2=80,33$	$\overline{X}_3=84,60$	$\overline{X}_4=69,80$

• Ipotezele statistice:

$$(H_0):\ m_1 = m_2 = m_3 = m_4$$

diferenţele dintre eşantioane sunt nesemnificative şi pot fi atribuite întâmplării (grupele sunt omogene, provin din aceeaşi populaţie statistică sub aspectul timpului de reacţie – efectul dozei de tranchilizant este nesemnificativ)

$$(H_1):\ \exists\ m_i \neq m_j$$ test F bilateral

diferenţele dintre eşantioane sunt semnificative (cel puţin două eşantioane provin din populaţii statistice diferite în privinţa variabilei studiate, deci doza de tranchilizant are un efect semnificativ asupra timupului de reacţie)

• Se calculează variabilitatea totală:

$$V_{tot} = \left(\sum x_1^2 + \sum x_2^2 + \ldots\right) - \frac{(T_1 + T_2 + \ldots)^2}{N_1 + N_2 + \ldots} = 939,59$$

Variabilitatea intergrupe

$$V_A = \left(\frac{T_1^2}{N_1} + \frac{T_2^2}{N_2} + \ldots\right) - \frac{(T_1 + T_2 + \ldots)^2}{N_1 + N_2 + \ldots} = 674,42$$

Variabilitatea intragrupe (eroare):

$$V_E = \left(\sum x_1^2 + \sum x_2^2 + \ldots\right) - \left(\frac{T_1^2}{N_1} + \frac{T_2^2}{N_2} + \ldots\right) = 265,17$$

Se verifică egalitatea $V_{tot} = V_A + V_E$

Scanned with OKEN Scanner

Se calculează estimatorii

$$S_A = \frac{V_A}{k-1} = \frac{674,42}{4-1} = 224,81 \quad S_E = \frac{V_E}{N-k} = \frac{674,42}{22-4} = 14,73 ,$$

unde $N = N_1 + N_2 + ... = 22$ şi k este numărul de eşantioane (4).

Valoarea calculată a parametrului F:

$$F_{calc} = \frac{S_A}{S_E} = \frac{224,81}{14,73} = 15,26$$

- Valoarea critică:

$$F_{cr}\Big|_{\substack{\nu_1=k-1 \text{ gr. lib.} \\ \nu_2=N-k \text{ gr. lib.} \\ \alpha=0,05}} = F_{cr}\Big|_{\substack{\nu_1=3 \text{ gr. lib.} \\ \nu_2=18 \text{ gr. lib.} \\ \alpha=0,05}} = 3,16$$

- $F_{calc} > F_{cr} \Rightarrow \begin{cases} \text{respingem } H_0 \\ \text{acceptam } H_1 \end{cases}$ deci doza de tranchilizant influenţează semnificativ timpul de reacţie.

- Raportul $V_A / V_{tot} = 674,42 / 939,59 = 0,718$ indică mărimea efectului variabilei studiate asupra rezultatelor. În cazul nostru, 71,8 % din variaţia timpului de reacţie se datorează variabilei independente (dozei de tranchilizant).

Analiza de variantă — bifactorială

4.2 Se studiază efectul oboselii asupra învăţării, în funcţie şi de natura materialului de reţinut. Gradul de oboseală se cuantifică prin 3 categorii — mic, mediu, mare — , iar materialul de învăţat poate fi verbal sau figural. Funcţie de valorile posibile ale celor două variabile, există 6 grupuri (eşantioane) de subiecţi ale căror rezultate la proba de reproducere sunt prezentate în tabelul următor. Studiaţi influenţele celor două variabile asupra rezultatelor, precum şi interacţiunea dintre acestea. (α =0,05)

Scanned with OKEN Scanner

	Material figural (B1)	Material verbal (B2)	Total A
Gr. ob. mic (A1)	7,9,8,10,10,7,8 $N_{A_1B_1}=7$ $T_{A_1B_1}=59$		
Gr. ob. mediu (A2)			
Gr. ob. mare (A3)			
Total B			

- Ipotezele statistice :

  (H_{0A}) : la nivelul populației variabila *A* (gradul de oboseală) nu influențează semnificativ rezultele învățării

  (H_{0B}) : la nivelul populației variabila *B* (natura materialului de reținut) nu influențează semnificativ rezultele învățării

  (H_{0AB}) : la nivelul populației nu există o interacțiune semnificativă între variabila *A* şi variabila *B*.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 7

REPARTIȚII ȘI TEOREME LIMITĂ (TL)

7.1. Considerații generale

Legile de repartiție clasice, aparent, sunt nelegate între ele. În realitate, între ele există conexiuni, unele dintre ele tind, la limită sau asimptotic, către altele. Acestea din urmă se numesc *legi de repartiție limită* sau repartiții limită.

Teoremele care justifică aceste limite se numesc teoreme limită.

În majoritatea cazurilor, dar nu totdeauna, *repartiția limită este repartiția normală.*

1. *Teorema Lindeberg-Levy*

Fie un şir de variabile aleatoare $x_1, x_2, ..., x_n$ independente şi care urmează aceeaşi repartiție. Se admit mediile $m = M(X_k)$ şi dispersiile $\sigma^2 = D^2(X_k)$ pentru $(\forall)$ $k = \overline{1,n}$.

Fie variabila aleatoare

$$Y_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_k - m) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - m}{\sigma} \qquad (7.1)$$

şi $F_n(x) = P(Y_n \langle x)$ funcția de repartiție a lui Y_n.

Atunci : $$\lim_{n\to\infty} F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

Observație. În condițiile teoremei Lindeberg-Levy

$Y_n \xrightarrow[n\to\infty]{} X \in N(0,1)$ indiferent de legea de repartiție.

Observație. *Repartiția normală standard*, notată $N(0,1)$, este un caz particular al repartiției normale generale $N(m,\sigma)$.

Observație. Se spune că variabila aleatoare X, continuă, urmează o lege de repartiție normală de parametri m şi σ dacă are forma :

$$X:\begin{pmatrix} x \\ f_{(x)} \end{pmatrix}, \quad f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}, \quad x\in R,\sigma\rangle \qquad (7.2)$$

Observații. Repartiția normală de parametri m și σ se mai notează cu $N(m,\sigma)$ sau $N(m,\sigma^2)$.

- Când se scrie $X \in N(m,\sigma)$ înseamnă că X urmează o repartiție normală.
- Legea normală mai poartă numele Gauss-Laplace, ei fiind cei dintâi care au utilizat-o în studiul problemelor de teoria erorilor.
- O reprezentare grafică (orientativă) a densității de repartiție normală poartă denumirea de „clopotul lui Gauss" cu vârful

$$v\left(x_v=m, \qquad y_v-\frac{1}{\sigma\sqrt{2\pi}}\right).$$

Grafice

Curba este simetrică față de dreapta $x=m$.

Efectuând schimbarea de variabilă $\frac{x-m}{\sigma}=y$, avem :

$$\int_{-\infty}^{\infty} f(x)dx = 1 = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-\frac{y^2}{2}\right)dy$$

$$M(x)=m; \qquad M(x^2)=m^2+\sigma^2; \qquad D^2(x)=\sigma^2; \qquad D(x)=\sigma$$

Definiție. Funcția $\phi: R \to R$ dată de relația :

$$\phi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z \exp\left(-\frac{t^2}{2}\right)dt \qquad (7.3)$$

se numește *funcția integrală a lui Laplace*, cu proprietățile :

$$\phi(0)=0$$

$$\phi(-z)=-\phi(z)$$

$$\phi(\infty)=\frac{1}{2}$$

$$\phi(-\infty)=-\phi(\infty)=-\frac{1}{2}$$

Scanned with OKEN Scanner

Dacă $X \in M(m,\sigma)$ atunci $F(X) = P(X\langle x) = \frac{1}{2} + \phi\left(\frac{x-m}{\sigma}\right)$ (7.4)

2. Teorema lui Liapunov

Fie $(X_n)_{n\geq 1}$ un şir de variabile aleatoare independente, astfel că $M(X_k) = M_k$, $D^2(X_k) = D_k^2$ şi $H_k^3 = M\left(\left|X_k - M_k\right|^3\right)$ există $(\forall)k \geq 1$.

Notăm: $S_n^2 = \sum_{k=1}^{n} D_k^2$; $K_n^3 = \sum_{k=1}^{n} H_k^3$; $Y_n = \frac{1}{S_n}\left(\sum_{1}^{n} X_k - \sum_{1}^{n} H_k\right)$.

Fie, de asemenea, $F_n(x) = P(Y_n\langle x)$ funcţia de repartiţie a variabilei aleatoare Y_n. În aceste condiţii, dacă $\lim_{n\to\infty}\frac{K_n}{S_n} = 0$ atunci:

$$\lim_{n\to\infty} F_n(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du \qquad (7.5)$$

- Remarcabile au fost rezultatele date de Moivre (1667–1754) şi Laplace (1749–1827) prin care se demonstrează că după un număr de probe suficient de mare repartiţia binomială poate fi aproximată prin repartiţia normală.
- Fie un experiment la a cărui efectuare poate să apară evenimentul A cu probabilitatea $P(A)=p$, sau contrariul său $\overline{A}$ cu

$$P(\overline{A}) = q = 1 - p.$$

Efectuându-se n probe, atunci variabila aleatoare X_n care reprezintă numărul de realizări ale lui A din cele n probe va fi:

$$X_n : \begin{pmatrix} k \\ p(k) \end{pmatrix}, \quad P(k) = p(n;k) = C_n^k p^k q^{n-k}, \quad 0 \leq k \leq n,$$

adică repartiţia binomială, având $M(X_n) = np$ şi $D^2(X_n) = npq$.

Conform teoremelor Moivre-Laplace, pentru n suficient de mare

$$X_n \in N\left(np, \sqrt{npq}\right).$$

Scanned with OKEN Scanner

TEOREMA integrală Moivre-Laplace

Dacă X_n este variabila aleatoare binominală, pentru n suficient de mare

$$P(X_n < x) = \frac{1}{\sqrt{2\pi npq}} \int_{-\infty}^{x} \exp\left(-\frac{(u-np)^2}{2npq} \right) du \qquad (7.6)$$

TEOREMA locală Moivre-Laplace

În condițiile teoremei integrale Moivre-Laplace

$$P(X_n = K) \cong \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(K-np)^2}{2npq}} \qquad (7.7)$$

7.2. Proprietățile distribuției binominale și distribuției normale

Distribuția binomială
(Distribuția lui Bernoulli)

Notăm cu „p" = probabilitatea de a se produce *un eveniment* („succes")
q = probabilitatea ca evenimentul să nu se producă („insucces")
p+q = 1

Probabilitatea ca un eveniment să se producă de x ori din „n" cazuri posibile (adică evenimentul are „succes" de x ori și n–x ori insucces) este dată de relația

$$p(x) = C_n^x p^x \cdot q^{n-x} = \frac{n!}{x!(n-x)!} p^x \cdot q^{n-x} \qquad (7.8)$$

Distribuția probabilității discrete din relația (7.8) se numește distribuție binominală

$$X : \begin{pmatrix} 0 & 1 & 2 & \dots & n \\ p(0) & p(1) & p(2) & & p(n) \end{pmatrix}$$

$$p(0) = C_n^0 p^0 q^n = 1 \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{0}{2}\right)^n = \frac{1}{2^n} \qquad p = \frac{1}{2} \quad q = \frac{1}{2}$$

$$p(1) = C_n^1 \left(\frac{1}{2}\right)^1 \cdot \left(\frac{1}{2}\right)^{n-1} = n \cdot \frac{1}{2^n}$$

Scanned with OKEN Scanner

Proprietăți ale distribuției binomiale

Media $\bar{x} = n\,p$

Dispersia $\nabla^2 = n\,p\,q$

Abater=ea pătratică medie $\nabla = \sqrt{n\,p\,q}$

Distribuția normală

Este una dintre cele mai importante distribuții de probabilitate continuă şi este cunoscută sub denumirea „curba normală“ sau „curba lui Gauss“ sau distribuția Gaussiană.

Este definită prin ecuația:

$$y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-\bar{x})^2}{\nabla^2}}$$

unde $\bar{x}$ = media; ∇ – abaterea pătratică medie

$\pi = 3{,}14159$

$e = 2{,}71828$

Variabilei x → **unitatea standard** $z = \dfrac{x - \bar{x}}{\nabla}$

îi corespunde sau **(valoarea standard)**

$$\Rightarrow \quad \underbrace{y = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}}$$

reprezintă forma standardizată a distribuției gaussiene cu media zero şi dispersia 1

Scanned with OKEN Scanner

- Aria de sub grafic este egală cu 1
- Aria de sub grafic este egală cu 1
- Între $z_1 = -1$ şi $z_2 = +1$ este o suprafață de 68,27% din Aria totală de 1 (100%)
- Între $z_1 = -2$ şi $z_2 = +2$ este o suprafață de 95,45%
- Între $z_1 = -3$ şi $z_1 = 3$ este o suprafață de 99,73%

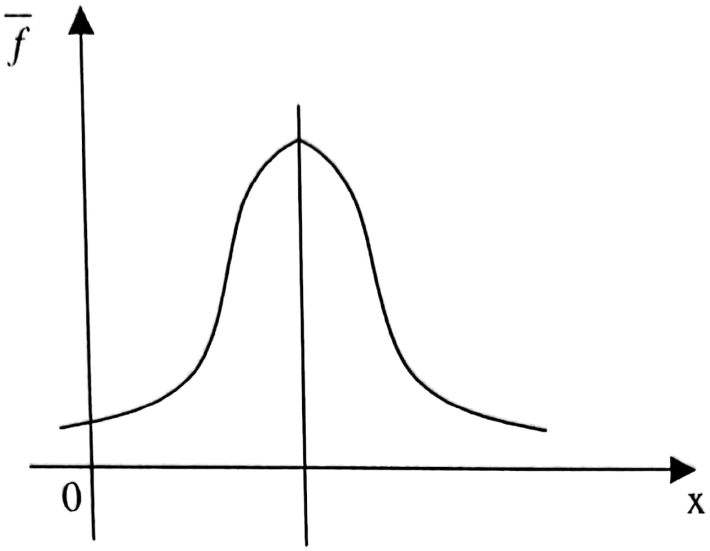
Distribuția normală se caracterizează prin aceea *că valorile caracteristicii, situate de o parte şi de alta a grupei cu frecvența cea mai mare* sunt aproximativ egale.

Un grafic în formă de clopot, mai alungit sau mai turtit, cu o tendință netă de simetrie sugerează forma distribuției normale.

În distribuțiile perfect simetrice:

- media
- mediana şi
- modul coincid

Scanned with OKEN Scanner

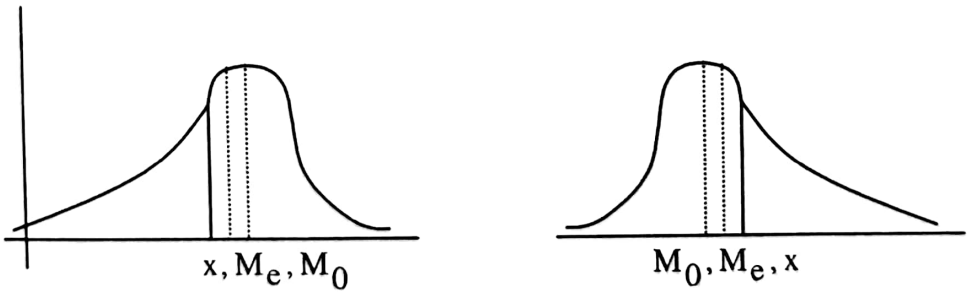


De menționat că simetria distribuției, precum şi *egalitatea* principalilor indicatori statistici (media, mediana şi modul) reprezintă *numai o condiție necesară, dar nu şi suficientă* pentru *a testa normalitatea distribuției*. Curba normală (sau clopotul lui Gauss) constituie reprezentarea funcției.

$$y = e^{-\frac{\left(\frac{x-m}{\sigma}\right)^2}{2}}$$

Distribuții asimetrice

Sunt acele distribuții în care se înregistrează o anumită polarizare a frecvențelor, fie în partea dreaptă, fie în partea stângă a clasei centrale. Când Me > $\overline{x}$ (media mai mică decât mediana) distribuția prezintă o asimetrie negativă. În cazul $\overline{x}$ > Me, asimetria distribuției este pozitivă.



Asimetrie negativă Asimetrie pozitivă

Deoarece media tinde să fie împinsă în direcția valorilor extreme ale „cozii" celei mai lungi a unei distribuții, asimetria poate fi calculată prin coeficienții:

Scanned with OKEN Scanner

- coeficientul de asimetrie raportat la mediană:

$$K = \frac{3(\bar{x} - Me)}{D}, \qquad -3 \le K \le +3$$

- coeficientul de asimetrie raportat la mod:

$$K = \frac{\bar{x} - M_0}{D}, \qquad -1 \le K \le +1$$

Cu cât valoarea coeficientului este mai aproape de zero cu atât asimetria este mai redusă, în timp ce valori apropiate de extreme reflectă asimetrii pronunțate.

Observație:

Distribuția teoretică se estimează cu ajutorul unei *distribuții de frecvențe* obținută dintr-un eșantion din populația cercetată.

Metoda generală utilizată constă în utilizarea mediei și abaterii pătratice medii a unui eșantion pentru aproximarea mediei și abaterii pătratice medii a populației originare. La sfârșit se probează exactitatea estimării distribuției teoretice utilizând testul x^2.

7.3. **Aplicații**

1. Probabilitatea de promovare a unui examen este de 0,75.

Determinați probabilitatea ca din 20 studenți să promoveze:

a. 15 studenți

b. 20 studenți $\qquad P_{(x)} = C_n^x \, p^x q^{n-x}$

Rezolvare: $p = 0{,}75; \qquad q = 0{,}25; \qquad n = 20; \qquad x = 10; 15$

a) P {10 studenți să promoveze} $= C_{20}^{10} 0{,}75^{10} \cdot 0{,}25^{10} =$

$$= C_{20}^{10} \left(\frac{75}{100}\right)^{10} \cdot \left(\frac{25}{100}\right)^{10} = C_{20}^{10} \left(\frac{3}{4}\right)^{10} \left(\frac{1}{4}\right)^{10} = \frac{20!}{10!\,10!} \frac{3^{10}}{4^{20}} =$$

$$= \frac{1 \cdot 2 \cdot 3 \ldots 10 \cdot 11 \cdot 12 \ldots 20}{1 \ldots 10 \cdot 1 \ldots 10} \frac{3^{10}}{2^{40}} = \frac{11 \cdot \overset{4}{12} \cdot 13 \cdot \overset{7}{14} \cdot \overset{3}{15} \cdot \overset{4}{16} \cdot 17 \cdot \overset{3}{18} \cdot 19 \cdot \overset{2}{20}}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10} \cdot \frac{3^{10}}{2^{40}} =$$

$$= \frac{11 \cdot \overset{2}{2} \cdot 13 \cdot \overset{2}{3} \cdot \overset{3}{2} \cdot 17 \cdot 19 \cdot 3^{10}}{\overset{3}{2} \cdot \overset{2}{3} \cdot 2^{40}}$$

Scanned with OKEN Scanner

$$P = 11 \cdot 13 \cdot 17 \cdot 19 \cdot \frac{3^{10}}{2^{38}} = \frac{46189 \cdot 59049}{1048576 \cdot 262144} = 0{,}009922$$

b) P {15 studenți să promoveze ex.}= $C_{20}^{15}\left(\frac{3}{4}\right)^{15}\left(\frac{1}{4}\right)^{5} =$

$$= \frac{15504}{2^{40}} \cdot 3^{15} = 15{,}140625 \cdot \left(\frac{3}{4}\right)^{15} = 0{,}2023287 \Leftrightarrow 20{,}23\%$$

2. Probabilitatea de faliment a unor firme este de 0,1 (p = 0,1; q = 0,9). Pentru un grup de 400 firme determinați:

a. media $\left(\bar{x} = np\right)$

b. dispersia $\left(\nabla^2 = n\,p\,q\right)$

c. abaterea pătratică medie $\left(\nabla = \sqrt{\nabla^2} = \sqrt{npq}\right)$

$n = 400$

Răspuns:

a. $\bar{x} = 400 \cdot 0{,}1 = 40$ *firme*

b. $\sigma^2 = npq = 400 \cdot 0{,}1 \cdot 0{,}9 = 36$

c. $\nabla = \sqrt{36} = 6$

3. Cifra de afaceri medie a unui grup de firme este de 720 mil. lei şi abaterea pătratică medie de 150 mil. lei.

Calculați valorile standard pentru următoarele cifre de afaceri:

a. $x_1 = 650$ *mil.lei* $\qquad \bar{x} = 720 \cdot 10^6$

b. $x_2 = 1100$ *mil.lei* $\qquad \nabla = 150 \cdot 10^6$

Răspuns:

$$Z_1 = \frac{x_1 - \bar{x}}{\nabla} = \frac{(650 - 720) \cdot 10^6}{150 \cdot 10^6} = \frac{-70}{150} = -0{,}46667$$

$$Z_2 = \frac{(1100 - 720) \cdot 10^6}{150 \cdot 10^6} = \frac{380}{150} = 2{,}5334$$

4. Cifra de afaceri medie a unui grup de firme este 800 mil.lei şi abaterea pătratică medie 120 mil.lei.

Scanned with OKEN Scanner

Calculați cifrele de afaceri a două firme cu valorile standard $Z_1 = -1{,}1$ și $Z_2 = 1{,}5$

Răspuns:

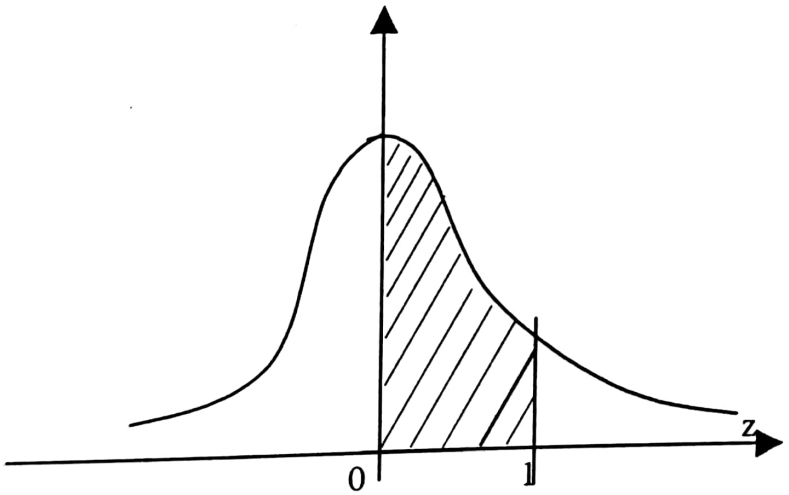
$$Z_1 = \frac{x_1 - \bar{x}}{\nabla};$$

$$x_1 = \bar{x} + Z_1 \nabla = [800 + (-1{,}1) \cdot 120]10^6 = 668 \text{ mil. lei}$$

$$Z_2 = x + Z_2 \cdot \nabla = (800 + 1{,}5 \cdot 120) \cdot 10^6 = 980 \text{ mil.lei}$$

Probabilitatea: $P\ (0 \le z \le 1) = 0{,}3413$

Această $P\ (0 \le z \le 1)$ = aria cuprinsă între z = 0, z = 1 de sub graficul curbei lui Gauss (graficul distribuției normale)



- Prin simetrie „Aria între z = 1 și z = 0" este egală cu aria cuprinsă între z = 0 și z = 1

$$P\{-0{,}65 \le z \le 0\} = P\{0 \le z \le 0{,}65\} = 0{,}2422$$

- $P\{-0{,}65 \le z \le 0{,}65\} = P\{-0{,}65 \le z \le 0\} + P\{0 \le z \le 0{,}65\} =$

$$= 2 \cdot P\{0 \le z \le 0{,}65\}$$

$$P\{0{,}83 \le z \le 1{,}92\} = P\{-0 \le z \le 1{,}92\} - P\{0 \le z \le 0{,}83\} =$$

$$= 0{,}4726 - 0{,}2967 = 0{,}1759$$

- $P\{-\infty \le z \le 0\} = 0{,}5$
- $P\{-\infty \le z \le -0{,}7\} = 0{,}5 - P\{0 \le z \le 0{,}7\} = 0{,}5 - 0{,}2580 = 0{,}2420$ (rezultatul reprezintă aria de la stânga lui — 0,7)

Scanned with OKEN Scanner

5. Dacă cifra de afaceri (CA) medie pentru 700 de firme este 180 mil.lei şi abaterea pătratică medie este 18 mil. lei, iar distribuția CA se consideră normală, determinați câte firme au:

a. CA cuprinsă între 160 şi 190 mil.lei

b. CA mai mare de 250 mil.lei?

Rezolvare:

a. $z = \frac{x - \bar{x}}{\nabla}$;

$$x_1 = 160 \text{ mil.lei} \Rightarrow z_1 = \frac{160 - 180}{18} = -\frac{20}{18} = -1{,}1111$$

$$x_2 = 190 \text{ mil.lei} \Rightarrow z_2 = \frac{190 - 180}{18} = -\frac{10}{18} = -0{,}555555$$

$$P\{-1{,}1111 \le z \le 0{,}5555\} = P\{0 \le z \le 0{,}5555\} + P\{0 \le z \le 1{,}1111\} =$$

$$= 0{,}2088 + 0{,}3438 = 0{,}5526$$

$$P = \frac{n_1}{700}; \qquad n_1 = 700\, P = 700 \cdot 0{,}5526 = 386{,}8 \sim 387$$

b. $x_3 = 250\ mil.lei \rightarrow z_3 = \frac{250 - 180}{18} = \frac{70}{18} = 3{,}8888$

$$P\{z \ge 3{,}888\} = 0{,}5 - P\{0 \le z \le 3{,}888\} = 0{,}5 - 0{,}4999 = 0{,}0001$$

$$0{,}0001 = \frac{n_2}{700}; \quad n_2 = 700 \cdot 0{,}0001 = 0{,}07$$

Observații:

$x_4 = 220$ mil.lei

$$z_4 = \frac{220 - 180}{18} = \frac{40}{18} = 2{,}2222$$

$$P\{z \ge 2{,}2222\} = 0{,}5 - P\{0 \le z \le 2{,}2222\} = 0{,}5000 - 0{,}4868 = 0{,}0132$$

$$0{,}0132 = \frac{n_4}{700}; \qquad n_4 = 700 \cdot 0{,}0132 = 9{,}24 \sim 9 \text{ firme}$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 8

ANALIZA DE REGRESIE ŞI CORELAŢIE

Analiza sistemelor naturale sau sociale impune studierea legăturilor dintre variabilele definitorii. În cazul variabilelor statistice, legăturile pot fi studiate cu ajutorul *metodei de regresie şi corelaţie*.

În fundamentarea teoriei corelaţiei şi-au adus contribuţii, în special, K. Pearson (corelaţia pentru trei variabile), Weldon, Edgeworth, G.U. Yule (corelaţia multiplă), Spearman (corelaţia rangurilor), Mosteller şi Tukey (corelaţia robustă).

8.1. Probleme ale analizei de regresie şi corelaţie

Legăturile dintre variabile pot apare trei situaţii:

- *Legătură nulă*, când nu există nici o influenţă între variabilele considerate (de exemplu, între culoarea ochilor şi performanţa şcolară);
- *Legătură funcţională*, când modificarea unei variabile conduce la variaţia altei variabile într-o măsură ce rămâne constantă, indiferent de tipul şi locul de referinţă;
- *Legătură statistică* (stochastică), când modificarea unei variabile este rezultatul conjugat al influenţei mai multor variabile, manifestată în medie pe ansamblul unităţilor unei colectivităţi şi nu în fiecare caz în parte (de exemplu, între nivelul consumului şi cel al veniturilor unei populaţii).

În studiul a două sau mai multor variabile se poate evidenţia o relaţie de *dependenţă funcţională* sau o *dependenţă statistică*.

Scanned with OKEN Scanner

8.1.1. Noțiuni de bază

Regresie și corelație. *Noțiunea de regresie exprimă o legătură de tip statistic și anume regresia în medie* cu privire la influența unor variabile, iar *conceptul de corelație exprimă raporturi reciproce între anumite caracteristici.*

Covariația. *Exprimă variația simultană a două variabile între care există o dependență.*

Analiza de regresie. Este o metodă statistică care permite studierea și măsurarea relației care există între două sau mai multe variabile, precum și descoperirea legii sub forma legăturilor dintre variabile.

Analiza de corelație. Analiza de corelație este o metodă statistică prin care se *măsoară intensitatea legăturii* dintre variabile. După forma modelului de regresie, corelația poate fi *corelație simplă* sau *corelație multiplă*.

Variabilă rezultativă. Variabilă factorială. Variabilă reziduală. Variabila estimată se numește *variabilă rezultativă* (dependentă, endogenă) și se notează prin Y. Variabila care explică variația variabilei *Y* se numește *variabilă factorială* (independentă, exogenă) și se notează prin *X*. Variabila care exprimă influența tuturor factorilor neincluși în model *se numește reziduu* (eroare de modelare) și se notează prin *e*.

Model de regresie. Un model de regresie poate fi scris:

$$Y = f(X_1, X_2, \ldots, X_n) + e$$

în care:

Y — variabila rezultativă (dependentă);

$X_1, \ldots, X_n$ — variabile factoriale (independente);

e — variabila eroare care reprezintă influența tuturor factorilor neincluși în model, considerați ca „eroare" (reziduu) de modelare.

Tipuri de legături statistice. Legăturile statistice se diferențiază după diferite criterii, cum ar fi: *numărul variabilelor corelate, sensul legăturii, forma legăturii* etc.

După numărul variabilelor corelate, avem legături simple și legături multiple.

Legăturile simple exprimă variația variabilei rezultative (Y) în funcție de o singură variabilă factorială (X). *Legăturile multiple* exprimă dependența variabilei (Y) în funcție de variația simultană a mai multor variabile factoriale (X_1, X_2,...).

Legăturile directe exprimă modificarea lui (Y) în același sens cu variabila (X).

Legăturile inverse exprimă variația lui (Y) în sens contrar variației variabilei (X).

După forma legăturilor, deosebim: legături rectilinii, legături curbiliniii.

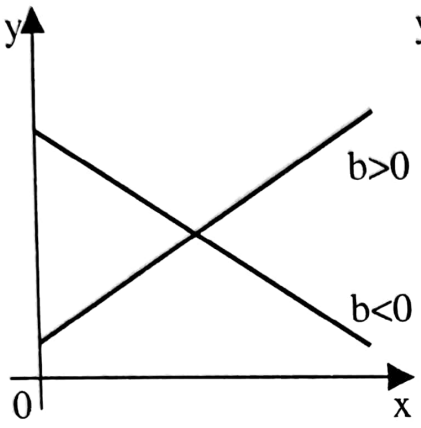
8.1.2. **Metode analitice de analiză a legăturilor statistice**

Metodele analitice permit exprimarea matematică a legăturii și determinarea numerică a intensității legăturii. Acestea sunt cunoscute sub *denumirea de modele de regresie și corelație*.

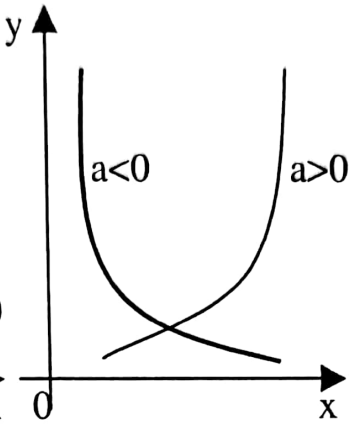
Modelele de regresie permit reprezentarea formei distribuției a două variabile corelate.

Cele mai utilizate modele de regresie în studiul fenomenelor și proceselor social economice sunt: a) liniar, b) exponențial, c) putere, d) parabolic, e) hiperbolic f) Tornqvist, g) Gompertz, h) logistic. Curbele corespunzătoare modelelor enumerate sunt prezentate în figura 8.1.

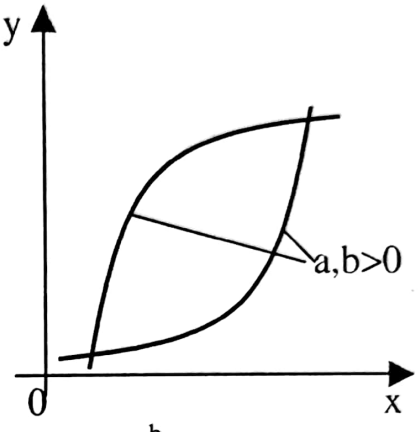
Scanned with OKEN Scanner



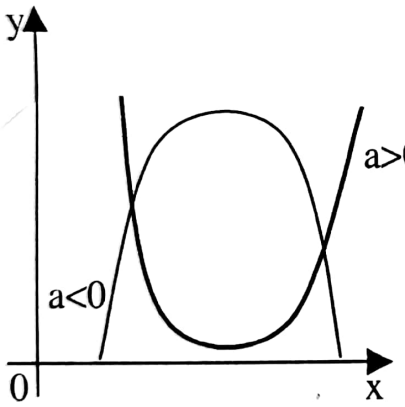
$y=a+bx$



$y=ab^x$

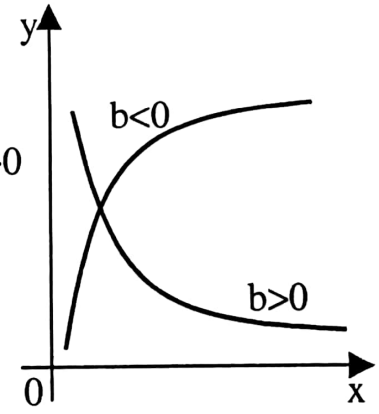


$y=ax^b$



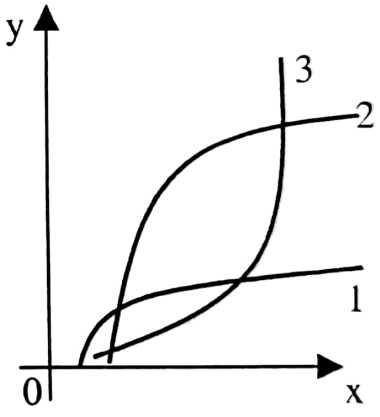
$y=a+bx+cx^2$

d)



$y=a+b\frac{1}{x}$

e)



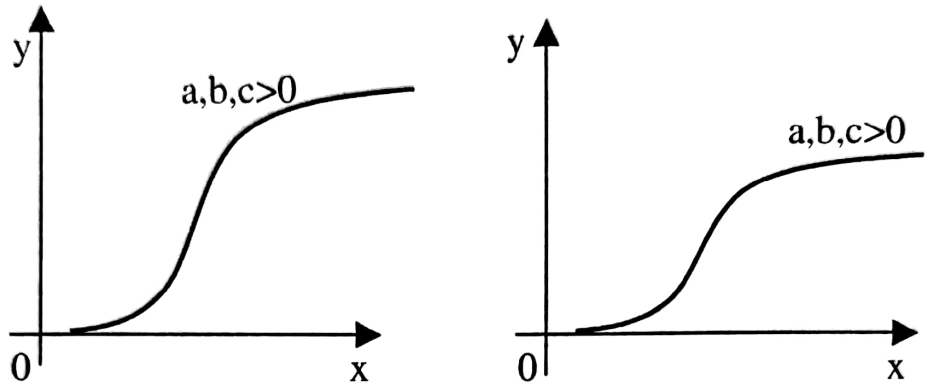
1) $y=\frac{x}{x+b}a$

2) $y=a\frac{x-c}{x+b}$

f)

Scanned with OKEN Scanner

Scanned with OKEN Scanner



g) h)

Fig. 8.1. *Modele de regresie*

8.2. **Modele de regresie cu două variabile**

Relația statistică între două variabile: *Y* (variabila dependentă) şi *X* (variabila independentă) se exprimă printr-o funcție de regresie:

$y_x = f(x) + e$

Alegerea modelului care să exprime legătura dintre variabilele corelate depinde în mare măsură de volumul datelor cunoscute.

8.2.1. **Modelul de regresie liniară**

Modelul de regresie liniară simplă exprimă relația dintre două variabile sub forma:

$$y_x = a + bx + e \qquad (8.1)$$

la nivelul unei populații şi

$$\hat{y}_x = \hat{a} + \hat{b}x + e \qquad (8.2)$$

la nivelul unui eşantion observat.

Funcțiile (8.1) şi (8.2) se numesc ecuații de regresie, a şi b sunt parametri necunoscuți ce urmează a fi estimați, iar $\hat{a}$ şi $\hat{b}$ sunt estimatorii parametrilor de regresie:

a — **ordonata la origine**, arată valoarea variabilei y când x = 0;
b — **panta dreptei**, numit şi coeficient de regresie;
e — variabila eroare aleatoare neobservabilă.

Semnul parametrului de regresie b indică direcţia legăturii dintre cel variabile corelate:

b > 0 indică o legătură directă (pozitivă);
b = 0 nu există legătură;
b < 0 legătură inversă (negativă).

Parametrii a şi b definesc dreapta şi sunt estimaţi cu datele observate pe un eşantion. Ca urmare, *analiza de regresie presupune estimarea parametrilor de regresie a şi b, alegând ca estimatori valorile â şi* $\hat{b}$. Se consideră acei estimatori pentru care valoarea variabilei aleatoare e_i, adică distanţa dintre o valoare reală y_i şi o valoare estimată, $y_{xi} = \hat{a} + \hat{b}\, x_i$, să fie cât mai mică.

De regulă, parametrii ecuaţiei de regresie se determină cu criteriul ($\Sigma(ei)^2$: minim), prin metoda celor mai mici pătrate:

$$\bar{S} = \Sigma\, e_i^2 = \Sigma\, (y_i - y_{xi})^2 = \text{minim} \qquad (8.3)$$

$$S = \Sigma\, (y_i - a - bx_i)^2 = \text{minim} \qquad (8.4)$$

Expresia S se minimizează anulând derivatele parţiale ale lui S în raport cu a şi b:

$\frac{\partial S}{\partial a} = 2\Sigma\, (y_i - a - bx_i\ )\,(-1) = 0$ $\frac{\partial S}{\partial a} = 2\Sigma\, (y_i - a - bx_i)\,(-x_i) = 0$ obţinem, un sistem de ecuaţii normale sub forma:

$$na + b\Sigma x_i = \Sigma y_i a$$
$$a\Sigma x_i + b\Sigma x^2{}_i = \Sigma x_i y_i \qquad (8.5)$$

Din sistemul de ecuaţii (8.5) se obţin parametrii a şi b:

$$b = \frac{\Delta b}{\Delta} = \frac{\Sigma\, x_i\, y_i - \Sigma\, x_i \Sigma\, y_i}{n\Sigma\, x_i^{\,2} - (\Sigma\, x_i)^2} \qquad (8.6)$$

Scanned with OKEN Scanner

$$a = \frac{\Delta a}{\Delta} = \frac{\Sigma y_i \Sigma x_i^2 - \Sigma x_i \Sigma x_i y_i}{n\Sigma x_1^2 - (\Sigma x_i)^2} \qquad (8.7)$$

$$a = \bar{y} - b\bar{x} \qquad (8.8)$$

8.2.2. **Modele de regresie neliniară**

Majoritatea legăturilor dintre variabilele, în domeniul economic, se încadrează în următoarele modele: parabolic, hiperbolic, exponențial, pe care le prezint în continuare. O altă parte se aduc la forma modelului liniar prin logaritmare şi se tratează similar cu modelele liniare.

A. Modelul de regresie de tip parabolic

Pentru exprimarea acestui model, se utilizează, de regulă, parabola de gradul doi:

$$y_x = a + bx + cx^2 \qquad (8.9)$$

Pentru determinarea parametrilor, aplicând metoda celor mai mici pătrate, obținem:

$$\begin{aligned} na + b\Sigma x_i + c\Sigma x^2{}_i &= \Sigma y_i \\ a\Sigma x_i + b\Sigma x^2{}_i + c\Sigma x^3{}_i &= \Sigma x_i y_i \\ a\Sigma x^2{}_i + b\Sigma x^3{}_i + c\Sigma x^4{}_i &= \Sigma x^2{}_i y_i \end{aligned} \qquad (8.10)$$

B. Modelul de regresie de tip hiperbolic: $y_x = a + \frac{b}{x}$ ai cărei parametri a şi b se determină din sistemul:

$$na + b\Sigma \frac{1}{x_i} = \Sigma y_i \qquad (8.11)$$

$$a\Sigma \frac{1}{x_i} + b\Sigma \frac{1}{x_i^2} = \Sigma \frac{y_i}{x_i}$$

C. Modelul de regresie de tip exponențial are forma:

$$y_x = a \cdot b^x \qquad (8.12)$$

Expresia (8.12) poate fi liniarizată prin logaritmare:

$$\log y_x = \log a + x \cdot \log b \qquad (8.13)$$

Aplicând metoda celor mai mici pătrate, se obține sistemul:

$$\begin{aligned} n \cdot \log a + \log b \cdot \Sigma x_i &= \Sigma \log y_i \\ \log a \cdot \Sigma x_i + \log b \cdot \Sigma x^2_i &= \Sigma x_i \cdot \log y_i \end{aligned} \qquad (8.14)$$

8.3. Măsurarea intensității legăturii dintre două variabile

După determinarea liniei de regresie, care exprimă legătura între variabile, se măsoară şi intensitatea legăturii. Aceasta arată gradul de concentrare sau de împrăştiere a valorilor y în jurul liniei de regresie y_x.

Intensitatea legăturii se exprimă cu ajutorul coeficientului de corelație şi a raportului de corelație. O apreciere aproximativă a intensității legăturii dintre variabile se poate obține prin aplicarea metodelor neparametrice.

8.3.1. Coeficientul de corelație

Coeficientul de corelație propus de K. Pearson se notează cu r_{y_x} şi este definit prin relația:

$$r_{y_x} = \frac{\text{cov}(x,y)}{S_x \cdot S_y} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n \cdot S_x \cdot S_y} \qquad (8.15)$$

în care:

cov(x,y) — covarianța;

x_i; y_i; şi $\bar{x}$; $\bar{y}$ — variabile corelate şi nivelul mediu al acestora;

n — numărul perechilor de valori;

S_x şi S_y — abaterea medie pătratică pentru x, respectiv y.

Analizând relația de calcul a coeficientului de regresie (8.6) şi cea a coeficientului de corelație (8.15) se constată că între aceşti indicatori există următoarea legătură:

Scanned with OKEN Scanner

$$r_{y_x} \cdot = b \cdot \frac{S_x}{S_y} \qquad (8.16)$$

din care rezultă că semnul coeficientului de corelație coincide cu semnul coeficientului de regresie, deoarece S_x și $S_y \geq 0$.

Valoarea coeficientului de corelație este cuprinsă între –1 și +1:

$-1 \leq r_{y_x} \cdot \leq +1$

Cele două valori extreme reprezintă o legătură liniară perfectă (funcțională între cele două variabile, „pozitivă", respectiv „negativă". Valoarea 0 semnifică absența legăturii între cele două variabile.)

8.3.2. **Raportul de corelație**

Este un indicator al intensității legăturii ce poate fi aplicat atât în cazul regresiei liniare cât și a celei neliniare simple sau multiple.

Raportul de corelație este notat cu η și se definește prin relația:

$$\eta = \sqrt{\frac{\sigma^2 y_x}{\sigma^2 y}} \qquad (8.17)$$

$$\text{sau } \eta = \sqrt{1 - \frac{\sigma^2 y/y_x}{\sigma^2 y}} \qquad (8.18)$$

în care:

$\sigma^2_y = \frac{\Sigma(y_i - \bar{y})^2}{n}$, reprezintă varianța generală, respectiv varianța variabilei y în raport cu media tuturor valorilor empirice:

$\sigma^2 y_x = \frac{\Sigma(y_x - \bar{y})^2}{n}$, varianța valorilor teoretice față de media lor (varianța sub influența factorilor esențiali);

Scanned with OKEN Scanner

$\sigma^2_{y/y_x} = \frac{\Sigma(y_i - y_x)^2}{n}$, varianța valorilor empirice față de valorile teoretice ale variabilei (varianța reziduală, eroare e_i).

Varianța generală este egală cu suma celorlalte două varianțe componente:

$$\sigma^2_y = \sigma^2_{y_x} + \sigma^2_{y/y_x},$$

respectiv, variația totală (V_T) este suma variației explicate (V_E) și a variației reziduale (V_R), adică: $V_T = V_E + V_R$.

8.4. **Corelația neparametrică**

Metodele neparametrice operează cu rangurile valorilor caracteristice. Pentru obținerea rangurilor unităților statistice se ordonează crescător sau descrescător, în funcție de variabila factorială.

Se vor prezenta următoarele metode neparametrice de studiere a corelației :

1. Metoda lui SPEARMAN.
2. Metoda lui KENDALL.
3. Metoda lui FECHNER.

Metoda lui SPEARMAN. Coeficient de corelație a rangurilor.

Fie două serii statistice X, Y cu valorile caracteristicilor x_i, y_i :

x_i	y_i
x_1	y_1
x_2	y_2
.	.
.	.
.	.
x_n	y_n

$\Rightarrow$ stabilim rangurile seriilor și diferența rangurilor d_i

r_{x_i}	r_{y_i}	$d_i = r_{x_i} - r_{y_i}$
r_{x_1}	r_{y_1}	d_1
r_{x_2}	r_{y_2}	d_2
.	.	.
.	.	.
.	.	.
r_{x_n}	r_{y_n}	d_n

Scanned with OKEN Scanner

În cazul unei legături funcţionale directe, între variabile va exista o concordanţă deplină între ranguri sau o disconcordanţă în cazul legăturilor funcţionale inverse.

Spearman a stabilit coeficientul θ :

$$\theta = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i}{n\left(n^2 - 1\right)}$$

denumit coeficientul de corelaţie a rangurilor SPEARMAN, unde $d_i = r_{x_i} - r_{y_i}$, n fiind numărul perechilor de valori corelate.

Coeficientul lui SPEARMAN este cuprins între –1 şi 1 şi rezultă din coeficientul de corelaţie al lui Pearson (r_{xy}) :

$$r_{xy} = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}_a\right)\left(y_i - \bar{y}_a\right)}{n \cdot \nabla_x \nabla_y}$$

unde : ∇_x, ∇_y sunt abaterile medii pătratice ale seriilor statistice X, Y.

$$\nabla_x = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_a\right)}{n}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}_a^2}\,,\quad \nabla_y = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}_a\right)}{n}} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}_a^2}$$

Ţinând seama că rangurile sunt cuprinse, pentru fiecare serie statistică, printre 1,2,3, ..., n şi că :

$$\bar{r}_x = \frac{1+2+\ldots+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2},$$

$$\bar{r}_y = \frac{1+2+\ldots+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

Scanned with OKEN Scanner

$$\theta = \frac{\sum_{i=1}^{n}\left(r_{x_i} - \bar{r}_x\right)\left(r_{y_i} - \bar{r}_y\right)}{n \cdot \sqrt{\frac{\sum_{i=1}^{n} r_{x_i}^2}{n} - \bar{r}_x^2} \cdot \sqrt{\frac{\sum_{i=1}^{n} r_{y_i}^2}{n} - \bar{r}_y^2}}$$

obținem în final $\theta = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2-1\right)}$.

În cazul unei legături funcționale directe, *există o concordanță deplină a rangurilor* $\theta = 1$, iar pentru legături funcționale inverse, $\theta = -1$.

r_{x_i}	r_{y_i}
1	1
2	2
3	3
.	.
.	.
.	.
n	n
(a)	

r_{x_i}	r_{y_i}
1	n
2	n – 1
3	n – 2
.	.
.	.
.	.
n	n – n + 1
(b)	

(a) $d_i = 0\ (\forall)\ i = \overline{1,n} \Rightarrow \theta = 1.$

(b) $d_i = r_{x_i} - r_{y_i} = \begin{bmatrix} 1-n \\ 3-n \\ 5-n \\ \dots \\ i-(n-i+1) = 2i-1-n \\ \dots \\ n-(n-n+1) = 2n-1-n \end{bmatrix}$

Scanned with OKEN Scanner

$$d_i^2 = \begin{bmatrix} (1-n)^2 = 1^2 - 2n + n^2 \\ (3-n)^2 = 3^2 - 2\cdot 3n + n^2 \\ \dots \\ (2i-1-n)^2 = (2i-1)^2 - 2(2i-1)n + n^2 \\ \dots \\ (2n-1-n)^2 = (2n-1)^2 - 2(2n-1)n + n^2 \end{bmatrix}$$

$$\sum_{i=1}^{n} d_i^2 = \left[1^2 + 3^2 + \dots (2n-1)^2\right] - 2n\left[1 + 3 + 5 + \dots + (2n-1)\right] + n^3$$

Observație.

$$1 + 3 + 5 + \dots + (2n-1) = \frac{(a_1 + a_n)n}{2} = \frac{(1 + 2n - 1)n}{2} = n^2$$

$$1^2 + 2^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

$$1^2 + 2^2 + \dots + n^2 + (n+1)^2 + (n+2)^2 + \dots + (2n)^2 = \frac{2n(2n+1)(4n+1)}{6} =$$

$$= \frac{n(2n+1)(4n+1)}{3}$$

$$\sum_{i=1}^{2n} i^2 = \sum_{i=1}^{n} i^2 + \sum_{i=n+1}^{2n} i^2$$

$$(n+1)^3 = n^3 + 3n^2 + 3n + 1$$

Scanned with OKEN Scanner

$$
\begin{array}{lrcccccc}
n=1 & 2^3 & = & 1^3 & + & 3\cdot 1^2 & + & 3\cdot 1 + 1 \\
n=2 & 3^3 & = & 2^3 & + & 3\cdot 2^2 & + & 3\cdot 2 + 1 \\
n=3 & 4^3 & = & 3^3 & + & 3\cdot 3^2 & + & 3\cdot 3 + 1 \\
\ldots & \ldots & & & & & & \\
n=n & (n+1)^3 & = & n^3 & + & 3n^2 & + & 3n + 1 \\
\hline
 & (n+1)^3 & = & 1 & + & 3S_2 & + & 3S_1 + n
\end{array}
$$

unde:

$$S_1 = \frac{n(n+1)}{2}$$

$$3S_2 = (n+1)^3 - \frac{3n(n+1)}{2} - (n+1)$$

$$3S_2 = \frac{n(n+1)(2n+1)}{2}$$

$$
\begin{array}{lrcccccc}
n=1 & 2^3 & = & 1^3 & + & 3\cdot 1^2 & + & 3\cdot 1 + 1 \\
n=2 & 3^3 & = & 2^3 & + & 3\cdot 2^3 & + & 3\cdot 2 + 1 \\
\ldots & \ldots & & & & & & \\
n=2n & (2n+1)^3 & = & (2n)^3 & + & 3\cdot(2n)^2 & + & 3\cdot(2n) + 1 \\
\hline
 & (2n+1)^3 & = & 3\left[1^2+2^2+\ldots+(2n)^2\right] & + & 3[1+2+\ldots+(2n)] & + & (2n+1)
\end{array}
$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

8.5. Corelaţie şi regresie (Extensii)

Regresie simplă liniară. Corelaţie liniară.

Fie două serii statistice X şi Y

v.c x_i	v.c y_i
x_1	y_1
x_2	y_2
.	.
.	.
.	.
x_n	y_n

Corelaţia indică ***gradul relaţiei dintre variabile***. În cazul a două variabile avem corelaţie simplă iar în *situaţia a trei sau mai multe variabile* avem corelaţie multiplă.

Regresia este o metodă analitică utilizată pentru calculul indicato-rilor sintetici ai legăturii dintre variabile, a intensităţii şi direcţiei legăturii.

Construim, la început, o diagramă a dispersiei punctelor $P_i(x_i, y_i)$ într-un sistem de coordonate carteziene. Dacă mulţimea punctelor P_i $(i=\overline{1,n})$ se grupează în jurul unei drepte

$$y = a + bx$$

avem **regresie liniară**.

Termenul de „*regresie*“ derivă din latină şi înseamnă „întoarcere“.

Dacă punctele $P_i(x_i, y_i)$ se grupează în jurul unei curbe avem regresie simplă curbilinie.

Funcţia de regresie poate fi:

$$y = a + bx + cx^2$$

4. Scriem că coordonatele punctelor P_1, P_2, P_3 verifică ecuaţia de regresie curbilinie:

$$\left\{\begin{array}{l} y_1 = a + b\,x_1 + c\,x_1^2 \\ y_2 = a + b\,x_2 + c\,x_2^2 \\ y_3 = a + b\,x_3 + c\,x_3^2 \end{array}\right\}$$

cu necunoscutele a, b, c. Ordonăm sistemul astfel:

$$\left\{\begin{array}{l} a + x_1 b + x_1^2 c = y_1 \\ a + x_2 b + x_2^2 c = y_2 \\ a + x_3 b + x_3^2 c = y_3 \end{array}\right\}$$

Aplicăm metoda lui Cramer:

Etapa 1. Calculăm determinantul coeficienților necunoscutelor.

$$\Delta = \begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix}$$ determinant Van der Monde x_1, x_2, x_3

$$\Delta = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$$

$$\Delta_a = \begin{vmatrix} y_1 & x_1 & x_1^2 \\ y_2 & x_2 & x_2^2 \\ y_3 & x_3 & x_3^2 \end{vmatrix}; \quad \Delta_b = \begin{vmatrix} 1 & y_1 & x_1^2 \\ 1 & y_2 & x_2^2 \\ 1 & y_3 & x_3^2 \end{vmatrix}; \quad \Delta_c = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}$$

$$a = \frac{\Delta_a}{\Delta}; \quad b = \frac{\Delta_b}{\Delta}; \quad c = \frac{\Delta_c}{\Delta}$$

Scanned with OKEN Scanner

$\Rightarrow$ $y = a + bx + cx^2$ = funcția de regresie curbilinie

Observație:

Metoda punctelor selecționate este subiectivă, de aceea se recomandă utilizarea metodei celor mai mici pătrate.

Determinarea parametrilor unei funcții de regresie se poate face prin două metode:

a) metoda punctelor selecționate

b) metoda celor mai mici pătrate

Metoda punctelor selecționate

- În cazul regresiei liniare se aleg două puncte situate pe o dreaptă trasată în dispersia punctelor (P_i) din sistemul cartezian, de regulă câte un punct la extremitățile dreptei.

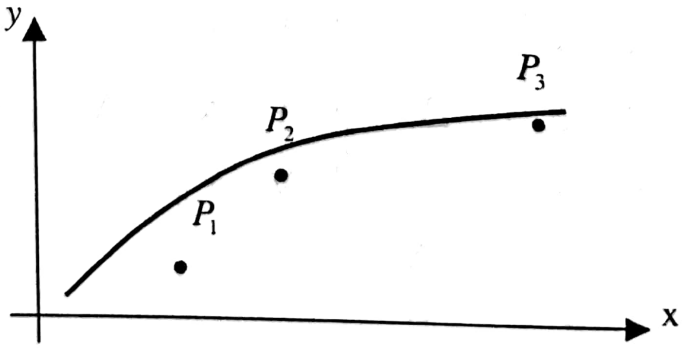
Fie $P_1(x_1, y_1)$ și $P_2(x_2, y_2)$ — punctele selecționate.

Scriem că P_1, P_2 se află pe dreapta $y = a + bx$:

$$\left.\begin{aligned} y_1 &= a + bx_1 \\ y_2 &= a + bx_2 \end{aligned}\right\} \Rightarrow a = \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}$$

$$b = \frac{y_2 - y_1}{x_2 - x_1}$$

- În cazul regresiei curbilinii de ordinul doi, cu funcția de regresie $y = a+bx+cx^2$, pentru determinarea parametrilor a, b, c se aleg trei puncte $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, $P_3(x_3, y_3)$, pe curba trasată vizual.



Rezultă:

$$a=\frac{\left(\sum_{i=1}^{n}x_i^2\right)\cdot\left(\sum_{i=1}^{n}y_i\right)-\left(\sum_{i=1}^{n}x_i\right)\cdot\left(\sum_{i=1}^{n}x_iy_i\right)}{n\left(\sum_{i=1}^{n}x_i^2\right)-\left(\sum_{i=1}^{n}x_i\right)^2}$$

$$b=\frac{n\left(\sum_{i=1}^{n}x_iy_i\right)-\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)}{n\left(\sum_{i=1}^{n}x_i^2\right)-\left(\sum_{i=1}^{n}x_i\right)^2}$$

Observații: În cazul seriilor cu frecvențe (serii cu ponderi n_i)

Serii	X	Y	Ponderi
	x_i	y_i	n_i
	x_1	y_1	n_1
	x_2	y_2	n_2
	.	.	.
	.	.	.
	.	.	.
	x_m	y_m	n_m

$$a=\frac{\left(\sum_{i=1}^{m}x_i^2n_i\right)\cdot\left(\sum_{i=1}^{m}y_in_i\right)-\left(\sum_{i=1}^{m}x_in_i\right)\left(\sum_{i=1}^{m}x_iy_in_i\right)}{\left(\sum_{i=1}^{m}n_i\right)\left(\sum_{i=1}^{m}x_i^2n_i\right)-\left(\sum_{i=1}^{m}x_in_i\right)^2}$$

$$b=\frac{\left(\sum_{i=1}^{m}n_i\right)\left(\sum_{i=1}^{m}x_iy_in_i\right)-\left(\sum_{i=1}^{m}x_in_i\right)\left(\sum_{i=1}^{m}y_in_i\right)}{\left(\sum_{i=1}^{m}n_i\right)\left(\sum_{i=1}^{m}x_i^2n_i\right)-\left(\sum_{i=1}^{m}x_in_i\right)^2}$$

Scanned with OKEN Scanner

Tabel de corelaţie de tipul (x_i, y_j, n_{ij})

y_j / x_i	y_1	y_2	y_3	. . .	y_n	n_j
x_1	n11	n12	n_{13}	. . .	n_{1n}	n_1
x	n21	n22	n_{23}	. . .	n_{2n}	n_2
.	.	.	.	. . .	.	.
.	.	.	.		.	.
.	.	.	.		.	.
x_n	n_{n1}	n_{n2}	n_{n3}	. . .	n_{nn}	n_n
TOTAL	n_1	n_2	n_3	. . .	n_n	$\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}$

$$\Rightarrow \begin{cases} n\,a + \left(\sum_{i=1}^{n} x_i\right) b + \left(\sum_{i=1}^{n} x_i^2\right) c = \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i\right) a + \left(\sum_{i=1}^{n} x_i^2\right) b + \left(\sum_{i=1}^{n} x_i^3\right) c = \sum_{i=1}^{n} x_i y_i \\ \left(\sum_{i=1}^{n} x_i^2\right) a + \left(\sum_{i=1}^{n} x_i^3\right) b + \left(\sum_{i=1}^{n} x_i^4\right) c = \sum_{i=1}^{n} x_i^2 y_i \end{cases}$$

din care se determină a,b,c şi apoi funcţia de regresie $y = a + bx + cx^2$.

- CAZUL (x_i, y_i, n_i)

x_i	y_i	n_i
x_1	y_1	n_1
x_2	y_2	n_2
.	.	.
.	.	.
.	.	.
x_m	y_m	n_m

Scanned with OKEN Scanner

Coeficienţii a, b, c se deduc din sistemul:

$$\begin{cases} \left(\sum_{i=1}^{m} n_i\right) a + \left(\sum_{i=1}^{m} x_i n_i\right) \cdot b + \left(\sum_{i=1}^{m} x_i^2 n_i\right) c = \sum_{i=1}^{m} y_i n_i \\ \left(\sum_{i=1}^{m} x_i n_i\right) a + \left(\sum_{i=1}^{m} x_i^2 n_i\right) b + \left(\sum_{i=1}^{m} x_i^3 n_i\right) c = \sum_{i=1}^{m} x_i y_i n_i \\ \left(\sum_{i=1}^{m} x_i^2 n_i\right) a + \left(\sum_{i=1}^{m} x_i^3 n_i\right) b + \left(\sum_{i=1}^{m} x_i^4 n_i\right) c = \sum_{i=1}^{m} x_i^2 y_i n_i \end{cases}$$

- În cazul tabelului de corelaţie parametrii a, b, c se obţin din sistemul:

$$\begin{cases} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} n_{ij}\right) a + \left(\sum_{i=1}^{n} x_i n_i\right) \cdot b + = \left(\sum_{i=1}^{n} x_i^2 n_i\right) c = \sum_{i=1}^{n} y_j n_j \\ \left(\sum_{i=1}^{n} x_i n_i\right) a + \left(\sum_{i=1}^{n} x_i^2 n_i\right) b + \left(\sum_{i=1}^{n} x_i^3 n_i\right) c = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_i n_{ij} \\ \left(\sum_{i=1}^{n} x_i^2 n_i\right) a + \left(\sum_{i=1}^{n} x_i^3 n_i\right) b + \left(\sum_{i=1}^{n} x_i^4 n_i\right) c = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^2 y_i n_{ij} \end{cases}$$

Metoda celor mai mici pătrate în cazul regresiei liniare

Metoda se bazează pe principiul:

„suma pătratelor diferenţelor dintre valorile empirice y_i şi cele teoretice $y_{x_i} = a + bx_i$ să fie minimă".

Notăm: $S = \sum_{i=1}^{n} (y_i - y_{x_i})^2 = \sum_{i=1}^{n} (y_i - a - b_{x_i})^2$

Punem condiţia S = minim, care este echivalentă cu condiţia de extrem pentru funcţii de două variabile S(a,b)

Scanned with OKEN Scanner

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases}$$

$$\Rightarrow \frac{\partial S}{\partial a} = 2\sum_{i=1}^{n}(y_i - a - bx_i)\cdot(-1) = 0 \quad \Leftrightarrow$$

$$\frac{\partial S}{\partial b} = 2\sum_{i=1}^{n}(y_i - a - bx_i)\cdot(-x_i) = 0$$

$$\sum_{i=1}^{n}(y_i - a - bx_i) = 0 \qquad (\sum_{i=1}^{n} y_i)\, na - b\,(\sum_{i=1}^{n} x_i) = 0$$

$$\Leftrightarrow \qquad \Leftrightarrow$$

$$\sum_{i=1}^{n}(y_i - a - bx_i)\,x_i = 0 \qquad \left(\sum_{i=1}^{n} x_i y_i\right) - a\left(\sum_{i=1}^{n} x_i\right) - b\sum_{i=1}^{n} x_i^2 = 0$$

Parametrii de regresie a şi b se obţin din sistemul

$$\begin{cases} n\,a + \left(\sum_{i=1}^{n} x_i\right) b = \sum_{i=1}^{n} y_i \\ \left(\sum_{i=1}^{n} x_i\right)\cdot a + \left(\sum_{i=1}^{n} x_i^2\right) b = \sum_{i=1}^{n} x_i y_i \end{cases}$$

$$a = \frac{\left(\sum_{i=1}^{n} x_i^2 n_i\right)\left(\sum_{j=1}^{n} y_j n_j\right) - \left(\sum_{i=1}^{n} x_i n_i\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j n_{ij}\right)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\right)\left(\sum_{i=1}^{n} x_i^2 n_i\right) - \left(\sum_{i=1}^{n} x_i n_i\right)^2}$$

Scanned with OKEN Scanner

$$b = \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j n_{ij}\right) - \left(\sum_{i=1}^{n} x_i n_i\right)\left(\sum_{j=1}^{n} y_j n_j\right)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\right)\left(\sum_{i=1}^{n} x_i^2 n_i\right) - \left(\sum_{i=1}^{n} x_i n_i\right)^2}$$

Regresie simplă curbilinie

Funcția de regresie $(x_i, y_i, 1)$

x_i	y_i
x_1	y_1
x_2	y_2
.	.
.	.
.	.
x_n	yn

$$y_{x_i} = a + bx_i + cx_i^2$$

$$S = \sum_{i=1}^{n}\left(y_i - y_{x_i}\right)^2 = \sum_{i=1}^{n}\left(y_i - a - bx_i - cx_1^2\right)^2$$

Condiție: $S = \min \Leftrightarrow \begin{cases} \frac{\partial S}{\partial a} = 0 \\ \frac{\partial S}{\partial b} = 0 \\ \frac{\partial S}{\partial c} = 0 \end{cases}$

$$\Leftrightarrow \begin{cases} 2\sum_{i=1}^{n}\left(y_i - a - bx_i - cx_i^2\right)(-1) = 0 \\ 2\sum_{i=1}^{n}\left(y_i - a - bx_i - cx_i^2\right)(-x_i) = 0 \\ 2\sum_{i=1}^{n}\left(y_i - a - bx_i - cx_i^2\right)(-x_i^2) = 0 \end{cases}$$

Scanned with OKEN Scanner

Prin ordonare obținem sistemul:

$$r = \frac{\sum_{i=1}^{n} \frac{x_i - \bar{x}}{\nabla_x} \cdot \frac{y_i - \bar{y}}{\nabla_y}}{n} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n\nabla_x \nabla_y} = \frac{\text{cov}(x, y)}{\nabla_x \nabla_y}$$

unde **covarianța** (cov (x,y) are expresia

$$\text{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Observații:

Coeficientul de corelație ia valori între — 1 şi 1

Dacă:

r = 1 — legătura dintre variabile este de tip funcțional, corelația este **pozitivă** (directă), perfectă.

r = –1 — legătura este de tip funcțional, iar între variabile corelația este **negativă** (inversă), perfectă

r = 0 — variabilele nu sunt corelate liniar (în acest caz trebuie căutată eventuala existență a unei corelații curbilinii)

- Variabilele sunt în general independente.

Observații: Coeficientul de corelație r se poate calcula şi cu relația

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}$$

Aplicație:

Să se determine prin metoda celor mai mici pătrate parametrii a,b,c din funcția de regresie $y=a+bx+cx^2$, pentru datele din următorul tabel:

Scanned with OKEN Scanner

Date inițiale

y =productivitatea muncii (mil. lei)		0,3–0,6	0,6–0,9	0,9–1,2	1,2–1,5	1,5–1,8	Total
x =salarii (mil, lei)	y_j / x_i	0,45	0,75	1,05	1,35	1,65	
0,4–0,6	0,5	4	–	–	–	–	4
0,6–0,8	0,7	5	6	–	–	–	11
0,8–1,0	0,9	1	8	3	2	–	14
1,0–1,2	1,1	–	–	1	4	2	7
1,2–1,4	1,3	–	–	–	–	4	4
TOTAL	–	10	14	4	6	6	40

Parametrii a, b, c se determină din sistemul (I)

$$(I)\begin{cases}\left(\sum_{i=1}^{5}\sum_{j=1}^{5} n_{ij}\right)a+\left(\sum_{i=1}^{5} x_i n_i\right)b+\left(\sum_{i=1}^{5} x_i^2 n_i\right)c=\sum_{j=1}^{5} y_j n_j\\ \left(\sum_{i=1}^{5} x_i n_i\right)a+\left(\sum_{i=1}^{5} x_i^2 n_i\right)b+\left(\sum_{i=1}^{5} x_i^3 n_i\right)c=\sum_{i=1}^{5}\sum_{j=1}^{5} x_i y_j n_{ij}\\ \left(\sum_{i=1}^{5} x_i^2 n_i\right)a+\left(\sum_{i=1}^{5} x_i^3 n_i\right)b+\left(\sum_{i=1}^{5} x_i^4 n_i\right)c=\sum_{i=1}^{5}\sum_{j=1}^{5} x_i^2 y_j n_{ij}\end{cases}$$

Scanned with OKEN Scanner

Date prelucrate

y_j / x_i	0,45	0,75	1,05	1,35	1,65	n_i	$x_i n$	$x_i^2 n_i$	$x_i^3 n_i$	$x_i^4 n_i$	$\sum_{j=1}^{5} x_i y_j n$	$\sum_{j=1}^{5} x_i^2 y_j$
0,5	4	–	–	–	–	4	2,0	1	0,5	0,25	0,9	0,45
0,7	5	6	–	–	–	11	7,7	5,39	3,773	2,6411	4,725	3,3075
0,9	1	8	3	2	–	14	12,6	11,34	10,206	9,1854	11,07	9,963
1,1	–	–	1	4	2	7	7,7	8,47	9,317	10,25	10,725	11,797 5
1,3	–	–	–	–	4	4	5,2	6,76	8,788	11,42	8,58	11,154
n_j	10	14	4	6	6	40	35,2	32,96	32,584	33,75	36	36,672
y_j / n_j	4,5	10,5	4,2	8,1	9,9	37,2	–	–	–	–	—	–

Coeficientul de corelaţie (r_{xy})

Arată intensitatea legăturii dintre variabilele corelate X şi Y.

Definiția 1.
Pentru serii statistice numim **abatere normată**

mărimea $Z_{x_i} = \frac{x_i - \bar{x}}{\nabla_x}$ sau $Z_{y_i} = \frac{y_i - \bar{y}}{\nabla_y}$

unde $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $\bar{y}$ $\frac{1}{n}\sum_{i=1}^{n} y_i$ = medii aritmetice

Scanned with OKEN Scanner

$$\text{sau}\left(\bar{x}=\frac{1}{n}\sum_{i=1}^{m}x_i n_i;\ \bar{y}=\frac{1}{n}\sum_{i=1}^{m}y_i n_i;\quad n=\sum_{i=1}^{m}n_i\right)$$

$$\nabla_x=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2};\quad \nabla_y=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2}=$$

= abateri standard sau abateri pătratice medii

$$\text{sau}\ \nabla_x=\sqrt{\frac{1}{n}\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2\cdot n_i};\quad \nabla_4=\sqrt{\frac{1}{n}\sum_{i=1}^{m}\left(x_i-\bar{x}\right)^2\cdot n_i}\qquad n=\sum_{i=1}^{m}n_i$$

Definiția 2

Coeficientul de corelație simplă liniară (r)

se calculează ca media aritmetică simplă a produsului abaterilor normate ale celor două variabile

cazul $(x_i, y_i, 1)$

$$r=r_{yx}=\frac{\sum_{i=1}^{n}z_{x_i}\cdot z_{y_i}}{n}$$

cazul (x_i, y_i, n_i) – serii statistice cu pondere

$$r_{xy}=\frac{\left(\sum_{i=1}^{m}n_i\right)\left(\sum_{i=1}^{m}x_i y_i n_i\right)-\left(\sum_{i=1}^{m}x_i n_i\right)\left(\sum_{i=1}^{m}y_i n_i\right)}{\sqrt{\left[\left(\sum_{i=1}^{m}n_i\right)\sum_{i=1}^{m}x_i^2 n_i-\left(\sum_{i=1}^{m}x_i n_i\right)^2\right]\left[\left(\sum_{i=1}^{m}n_i\right)\sum_{i=1}^{m}y_i^2 n_i-\left(\sum_{i=1}^{m}y_i n_i\right)^2\right]}}$$

- Dacă datele se prezintă sub forma unui tabel de corelație:

$$r_{xy}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_i-\bar{x}\right)\left(y_j-\bar{y}\right)n_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n}n_{ij}\nabla_x\nabla_y}$$

Scanned with OKEN Scanner

$$\text{cu } \nabla_x = \sqrt{\frac{\sum_{i=1}^{m}\left(x_i - \bar{x}\right)^2 n_i}{\sum_{i=1}^{m} n_i}}; \qquad \nabla_y = \sqrt{\frac{\sum_{i=1}^{m}\left(y_i - \bar{y}\right)^2 n_i}{\sum_{i=1}^{m} n_i}}$$

$$\Rightarrow$$

$$r_{xy} = \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j n_{ij}\right) - \left(\sum_{i=1}^{n} x_i n_i\right)\left(\sum_{j=1}^{n} y_j n_j\right)}{\sqrt{\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\right)\left(\sum_{i=1}^{n} x_i^2 n_i\right) - \left(\sum_{i=1}^{n} x_i n_i\right)^2\right]\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n} n_{ij}\right)\left(\sum_{i=1}^{n} y_i^2 n_i\right) - \left(\sum_{i=1}^{n} y_i n_i\right)^2\right]}}$$

$$\sum_{i=1}^{5}\sum_{j=1}^{5} x_i y_1 n_{ij} = \sum_{j=1}^{5} x_1 y_j n_{ij} + \sum_{j=1}^{5} x_2 y_j n_{2j} + \sum_{j=1}^{5} x_3 y_j n_{3j} + \sum_{j=1}^{5} x_4 y_j n_{4j} +$$

$$+ \sum_{j=1}^{5} x_5 y_j n_{5j} =$$

$$=(0{,}5\cdot\ 0{,}45\cdot\ 4)+(0{,}7\cdot\ 0{,}45\cdot\ 5+0{,}7\cdot\ 0{,}75\cdot\ 6)$$
$$+(0{,}9\cdot\ 0{,}45\cdot\ 1+0{,}9\cdot\ 0{,}75\cdot\ 8+0{,}9\cdot\ 1{,}05\cdot\ 3+0{,}9\cdot\ 1{,}35\cdot\ 2)+$$
$$+(1{,}1\cdot\ 1{,}05\cdot\ 1+1{,}1\cdot\ 1{,}35\cdot\ 4+1{,}1\cdot\ 1{,}65\cdot\ 2)+(1{,}3\cdot\ 1{,}65\cdot\ 4)$$

$$\sum_{i=1}^{5}\sum_{j=1}^{5} x_1^2 y_j n_{ij} = \sum_{j=1}^{5} x_1^2 y_j n_{ij} + \sum_{j=1}^{5} x_2^2 y_j n_{2j} + \sum_{j=1}^{5} x_3^2 y_j n_{3j} + \sum_{j=1}^{5} x_4^2 y_j n_{4j} +$$

$$+ \sum_{j=1}^{5} x_5^2 y_j n_{5j}$$

Sistemul (I) devine:

$$\begin{cases} 40a + 35{,}2b + 32{,}96c = 37{,}2 \\ 35{,}2a + 32{,}96b + 32{,}584c = 36 \\ 32{,}96a + 32{,}584b + 33{,}75\cdot c = 36{,}672 \end{cases}$$

Scanned with OKEN Scanner

$$\begin{pmatrix} 40 & 35{,}2 & 32{,}96 & 37{,}2 \\ 35{,}2 & 32{,}96 & 32{,}584 & 36 \\ 32{,}96 & 32{,}584 & 33{,}75 & 36{,}672 \end{pmatrix} \begin{matrix} \\ \cdot(-35{,}2) \\ \cdot(40) \end{matrix} \Big\} + \Big| \begin{matrix} \\ \cdot(-32{,}96) \\ \cdot(40) \end{matrix} \Big\} +$$

(a) (b)

$$\begin{pmatrix} 40 & 35{,}2 & 32{,}96 & 37{,}2 \\ 0 & +9{,}36 & 143{,}168 & 130{,}56 \\ 0 & 143{,}168 & 263{,}64 & 240{,}768 \end{pmatrix} \begin{matrix} \\ \cdot(143{,}168) \\ \cdot\underline{(-79{,}36)} \end{matrix} \Big\} +$$

$$\begin{pmatrix} 40 & 35{,}2 & 32{,}96 & 37{,}2 \\ 0 & 79{,}36 & +143{,}168 & \cdot 130{,}56 \\ 0 & 0 & -425{,}394 & -415{,}334 \end{pmatrix}$$

$$\begin{cases} 40a + 35{,}2b + 32{,}96c = 37{,}2 \\ 79{,}36b + 143{,}168c = 130{,}56 \\ 0 \quad 0 \quad - 425{,}394c = 415{,}334 \end{cases}$$

$$\Rightarrow c = \frac{415{,}334}{425{,}394} = 0{,}9763513$$

$$79{,}36\,b = 130{,}56 - 143{,}168 \cdot 0{,}9763513 =$$

$$= 130{,}56 - 139{,}78226 = -9{,}22226$$

$$b = -\frac{9{,}22226}{79{,}36} = 0{,}1162079$$

$$40\,a = 37{,}2 + 35{,}2 \cdot 0{,}1162079 - 32{,}96 \cdot 0{,}9763513 =$$

$$= 37{,}2 + 4{,}09 - 32{,}18 = 9{,}11; \qquad a = 0{,}22775$$

Scanned with OKEN Scanner

$$V_2 \quad (I) \Leftrightarrow \begin{cases} a + 0{,}88\,b + 0{,}824\,c = 0{,}93 \\ a + 0{,}936\,b + 0{,}926\,c = 1{,}0227 \\ a + 0{,}988\,b + 1{,}02397\,c = 1{,}1126 \end{cases}$$

$$\begin{cases} a + 0{,}88\,b + 0{,}824\,c = 0{,}93 \\ 0{,}056\,b + 0{,}102\,c = 0{,}0927 \\ 0{,}108\,b + 0{,}19997c = 0{,}1826 \end{cases}$$

$$\begin{cases} a + 0{,}88\,b + 0824\,c = 0{,}93 \\ b + 1{,}8214\,c = 1{,}655 \\ b + 1{,}8515\,c = 1{,}69107 \end{cases}$$

$$\begin{cases} a + 0{,}88\,b + 0{,}824\,c = 0{,}43 \\ b + 1{,}8214\,c = 1{,}655 \\ 0{,}0302\,c = 0{,}0357 \rightarrow c = 1{,}18313 \end{cases}$$

$$b = 1{,}655 - 1{,}8214 \cdot c = 1{,}655 - 2{,}1549 = -0{,}4999$$

V_3 Manual $a = 0{,}1382279$

$b = -0{,}0295858$

$c = 0{,}929292$

Ecuațiile curbelor de interpolare

Variabila **x se numește variabilă independentă** iar y= variabilă dependentă. Toate literele diferite de x și y sunt parametrii constanți.

Tipuri de curbe de interpelare

1) Dreaptă: $y = a_0 + a_1 x$

2) Parabolă: $y = a_0 + a_1 x + a_2 x^2$

3) Curbă cubică: $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$

Scanned with OKEN Scanner

4) Curbă de gradul IV: $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$

5) Curbă de gradul n: $y = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$

6) Hiperbolă $y = \dfrac{1}{a_0 + a_1 x}$ sau $\dfrac{1}{y} = a_0 + a_1 x$

7)Curbă exponenţială

$$y = a \cdot b^x \text{ sau } \lg y = (\lg a) + (\lg b) = a_0 + a_1 x$$

8) *Curbă geometrică* $y = a \cdot x^b$ sau $\lg y = \lg a + 1 \cdot \lg x$

9) *Curbă exponenţială modificată* $y = ab^x + g$

10) Curba lui GOMPERTZ

$$y = p \cdot q^{b^x} \text{ sau } \lg y = \lg p + b^x \cdot \lg q = ab^x + g$$

11) Curba lui GOMPERTZ modificată $y = p \cdot q^{b^x} + h$

12) Curba logistică: $y = \dfrac{1}{ab^x + g}$ sau $\dfrac{1}{j} = ab^x + g$

13) Curba semilogaritmică: $y = a_0 + a_2 \lg x + a_2 \lg^2 x$

Pentru a alege curba de interpolare se trasează diagrama dispersiei variabilelor transformate

$$(y \to \lg y; \quad x \to \lg x \quad \to \text{pentru curba } (8)).$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 9

SERII CRONOLOGICE

Analiza seriilor cronologice presupune caracterizarea variației în timp a unui fenomen. Se urmăreşte măsurarea creşterilor sau descreşterilor de nivel şi a modificărilor de structură, precum şi determinarea influenței factorilor sub incidența cărora are loc evoluția fenomenelor. Pentru a răspunde acestui obiectiv, statistica şi-a creat şi dezvoltat metode specifice pentru observarea şi analiza seriilor cronologice cum ar fi:

– metoda indicatorilor statistici ai dinamicii, în mărimi absolute, relative şi medii;

– metoda indicilor dinamicii fenomenelor, pentru a determina variația de la o perioadă la alta, precum şi influența factorilor;

– metode de analiză a componentelor, pentru a estima linia de tendință (trend) pentru măsurarea oscilațiilor sezoniere, pentru aprecierea variațiilor aleatoare;

– metode de prognoză statistică bazate pe extrapolarea trendului.

9.1. Indicatori pentru caracterizarea nivelului şi variației în timp

Seria cronologică, numită şi serie de timp, reprezintă un şir de observații numerice date în timp. În funcție de cum este definit timpul, prin momente, Ti, sau prin intervale, $t_i = T_i - T_{i-1}$, pot fi serii de timp de momente, respectiv, serii de timp de intervale.

Seriile de timp de momente sunt definite prin cuplurile de valori (T_i, y_i) şi prezintă volumul unei colectivități la diferite momente. Seriile de timp de intervale sunt definite prin cuplurile de valori (t_i, y_i) şi prezintă nivelul unui fenomen în diferite intervale de timp. Volumul acestuia se află prin însumarea termenilor seriei. De exemplu, producția obținută de o firmă într-un an.

Scanned with OKEN Scanner

9.1.1. **Indicatori absoluți**

Această categorie de indicatori exprimă, în unități concrete de măsură (kg. m, lei, etc.), nivelul sau modificările de nivel ale unui fenomen privit în evoluția sa în timp.

Nivelul absolut. Nivelul absolut reprezintă valoarea y_i a fiecărui termen al seriei cronologice.

Volumul absolut. Volumul se obține prin însumarea nivelurilor absolute (ale seriilor de intervale):

$$y = \sum_i y_i = y_1 + \ldots + y_i + \ldots + y_n \qquad (9.1)$$

Sporul absolut (modificarea absolută, creşterea sau descreşterea absolută).. Sporul absolut arată, în mărime absolută, cu cât s-a modificat, un fenomen, într-o perioadă sau moment față de o altă perioadă sau un alt moment considerat ca bază de comparare.

Sporul se poate determina fie în funcție de o bază mobilă, fie în funcție de o bază fixă.

Sporul cu bază fixă ($\Delta_{i/0}$) se calculează ca diferența între oricare termen (y_i) şi termenul inițial (y_0), după relația:

$$\Delta^y_{i/0} = y_i - y_0 \qquad (9.2)$$

Sporul cu bază mobilă, sau în lanț ($\Delta_{i/i-1}$)se calculează ca diferența între doi termeni consecutivi:

$$\Delta^y_{i/i-1} = y_i - y_{i-1} \qquad (9.3)$$

Relații între sporuri. Între sporurile cu bază fixă şi sporurile cu bază mobilă, calculate pentru aceeaşi serie, sunt valabile următoarele relații:

$$1.\ \sum_{i=0}^{n} \Delta^y_{i/i-1} = \Delta^y_{n/0} \qquad (9.4)$$

$$2.\ \Delta^y_{i/0} - \Delta^y_{i-1/0} = \Delta^y_{i/i-1} \qquad (9.5)$$

Scanned with OKEN Scanner

9.1.2. **Indicatori relativi**

În grupa de indicatori relativi se includ ritmul sau indicele de variație și ritmul sporului. Aceşti indicatori se obțin prin raportarea a două nivele comparate şi se exprimă sub formă de coeficient sau procente.

Ritmul sau indicele de variație. Ritmul de variație arată de câte ori a crescut sau scăzut nivelul unui fenomen într-o anumită perioadă (moment) față de nivelul aceluiaşi fenomen dintr-o altă perioadă (moment) considerată ca bază, şi se calculează după relațiile:

— ritmul cu bază fixă: $R^{y}_{i/0}=\frac{y_i}{y_0}$, respectiv $R^{y}_{i/0}=\frac{y_i}{y_0}\cdot 100$; (9.6)

— ritmul cu bază mobilă: $R^{y}_{i/i-1}=\frac{y_i}{y_{i-1}}$, respectiv $R^{y}_{i/i-1}=\frac{y_i}{y_{i-1}}\cdot 100$ (9.7)

Relații între ritmuri. Între ritmurile cu bază fixă şi ritmurile cu bază mobilă se stabilesc următoarele relații:

1. $\prod_{i=0}^{n} R_{i/i-1} = R_{n/0}$ (9.8)

2. $R_{i/0} \;/\; R_{i-1/0} = R_{i/i-1}$ (9.9)

Ritmul sporului. Acest indicator arată cu cât s-a modificat, în mărime relativă, nivelul fenomenului în perioada raportată față de nivelul din perioada de raportare. Ritmul sporului se calculează, în funcție de bază de raportare, fie cu bază fixă, fie cu bază mobilă, după relațiile:

— ritmul sporului cu bază fixă:

$$r^{y}_{i/0} = \frac{\Delta^{y}_{i/0}}{y_0} = \frac{y_i - y_0}{y_0} = \frac{y_i}{y_0} - 1 = R_{i/0} - 1 \qquad (9.10)$$

sau $r^{y}_{i/0} = \frac{\Delta^{y}_{i/0}}{y_0}\cdot 100 = \frac{y_i - y_0}{y_0}\cdot 100 = \frac{y_i}{y_0}\cdot 100 - 100$

— ritmul sporului cu bază mobilă:

$$r^{y}_{i/i-1} = \frac{\Delta^{y}_{i/i-1}}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} = \frac{y_i}{y_{i-1}} - 1 = R_{i/i-1} - 1 \qquad (9.11)$$

respectiv, $r^{y}_{i/i-1} = \frac{\Delta^{y}_{i/i-1}}{y_{i-1}} \cdot 100 = \frac{y_i}{y_{i-1}} \cdot 100 - 100 = R_{i/i-1} - 100$

Cu ajutorul sporurilor şi al ritmurilor sporului se poate afla valoarea absolută a unui procent de creştere, după relaţiile:

$\%(y_{i/0}) = \frac{\Delta_{i/0}}{r_{i/0}}$, $\%(y_{i/i-1}) = \frac{\Delta_{i/i-1}}{r_{i/i-1}}$, în care:

$\%(y_{i/0})$ — valoarea absolută a unui procent de creştere cu bază fixă;

$\%(y_{i/i-1})$ — valoarea absolută a unui procent de creştere cu bază mobilă.

9.1.3. **Indicatori medii**

Indicatorii medii se referă fie la nivelul, fie la variaţia termenilor unei serii cronologice. Variaţia medie poate fi exprimată fie în mărimi absolute (sporul mediu), fie în mărimi relative (ritmul sau indicele mediu de creştere).

Nivelul mediu. Pentru seriile cronologice nivelul mediu se calculează în funcţie de tipul seriei. Astfel, pentru o serie de timp de intervale, nivelul mediu se află calculând o medie aritmetică a termenilor seriei, iar pentru o serie de timp de momente, o medie cronologică.

Media cronologică simplă. Pentru o serie cronologică cu momentele egal distanţate între ele, media se calculează după relaţia:

$$\bar{y}_{cr} = \frac{\frac{y_1}{2} + y_2 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}, \qquad (9.12)$$

Scanned with OKEN Scanner

Media cronologică ponderată. Pentru o serie cronologică cu momentele inegal distanțate între ele, media se calculează cu relația:

$$\overline{y}_{cr} = \frac{\frac{y_1+y_2}{2}t_1 + \frac{y_2+y_3}{2}t_2 + \ldots + \frac{y_{n-1}+y_n}{2}t_{n-1}}{t_1+t_2+\ldots+t_{n-1}}, \qquad (9.13)$$

Media cronologică ponderată se mai poate calcula şi cu formula:

$$\overline{y}_{cr} = \frac{y_1\frac{t_1}{2} + y_2\frac{t+t_2}{2} + \ldots + y_n\frac{t_{n-1}}{2}}{t_1+t_2+\ldots+t_{n-1}}, \qquad (9.14)$$

unde t_i reprezintă mărimea intervalelor dintre două momente consecutive.

Sporul mediu ($\overline{\Delta}$). Sporul mediu reflectă modificarea (creşterea sau descreşterea) medie pe unitate de timp înregistrată de un fenomen într-o perioadă. Se calculează ca o medie a sporurilor individuale cu bază în lanț cu relația:

$$\overline{\Delta}_y = \frac{\sum_{i=0}^{n}\Delta_{i/i-1}}{n} = \frac{\Delta_{n/0}}{n} = \frac{y_n - y_0}{n} \qquad (9.15)$$

unde „n" reprezintă numărul sporurilor cu bază mobilă.

Ritmul mediu al variației ($\overline{R}$). Este numit şi indice mediu de variație şi arată de câte ori s-a modificat în medie pe an nivelul unui fenomen într-o perioadă în care fenomenul evoluează după o progresie geometrică.

Metoda mediei geometrice. Ritmul mediu al variației se calculează, de regulă, după metoda mediei geometrice, pe baza relației:

$$\overline{R} = \sqrt[n]{\prod_{i=0}^{n} R_{i/i-1}} \qquad (9.16)$$

în care n reprezintă numărul ritmurilor cu bază mobilă.

Scanned with OKEN Scanner

Ştiind că $\prod_{i=0}^{n} R_{i/i-1} = R_{n/0}$, relaţia (9.16) devine: $\overline{R} = \sqrt[n]{R_{n/0}} = \sqrt[n]{\frac{y_n}{y_0}}$

Folosind logaritmii, relaţia (9.16) devine: $\log \overline{R} = \frac{1}{n}\sum_{i=0}^{n} \log R_{i/i-1,}$ respectiv

$\log \overline{R} = \frac{1}{n} \log R_{n/0,}$ de unde $\overline{R}$ = antilog ($\log \overline{R}$).

Metoda mediei geometrice are dezavantajul că ţine seama numai de valoarea ultimului şi primului termen al seriei, ignorându-se termenii intermediari. Pentru a elimina acest dezavantaj se apelează la: metoda mediei parabolice propusă de (matematicianul român Froda) şi **metoda trendului exponenţial.**

Metoda mediei parabolice. În calculul ritmului mediu, după această metodă se consideră suma termenilor seriei şi termenul iniţial.

Fie o serie dinamică, $y_0, y_1, ..., y_n$, ai cărei termeni variază, aproximativ după o progresie geometrică. Plecând de la suma termenilor seriei:

$$y_0 + y_1 + ... + y_n = \sum_{i=0}^{n} y_i,$$

orice termen y_i al unei astfel de serii poate fi definit astfel:

$y_i = y_0 \cdot \overline{R^i}$; i = 1,2,..., n

În aceste condiţii, suma termenilor seriei poate fi scrisă astfel:

$$y_0 + y_0\overline{R} + y_0\overline{R^2} + ... + y_0\overline{R^n} = \sum_{i=0}^{n} y_i;\ y_0(1 + \overline{R} + \overline{R^2} + ... + \overline{R^n} = \sum_{i=0}^{n} y_i;$$

$$1 + \overline{R} + \overline{R^2} + ... + \overline{R^n} = \frac{\sum_{i=0}^{n} y_i}{y_0}$$

În primul membru al relaţiei anterioare avem suma termenilor unei progresii geometrice, cu raţia $\overline{R}$ şi primul termen egal cu 1.

Scanned with OKEN Scanner

$$1+\bar{R}+\bar{R}^2+\ldots+\bar{R}^n=\frac{\bar{R}^{n+1}-1}{\bar{R}-1},$$

prin urmare: $$\frac{\bar{R}^{n+1}-1}{\bar{R}-1}=\frac{\sum_{i=0}^{n} y_i}{y_0} \qquad (9.17)$$

Determinarea ritmului mediu pe baza acestei relații se face iterativ.

Metoda are dezavantajul că exagerează importanța primului termen, lucru care se elimină prin utilizarea metodei trendului exponențial ($y_t = a\bar{R}^t$;).

Metoda trendului exponențial. Calculul ritmului mediu pe baza metodei trendului exponențial presupune considerarea ritmului mediu egal cu parametrul de trend b al unei funcții de trend exponențiale, y_t=a· b^{ti}, adică: $\bar{R}$=b. Coeficientul de trend exponențial b, respectiv $\bar{R}$, se determină cu metoda celor mai mici pătrate (Σe_i^2=min.), după relația: $\bar{R}$ = anti log (log b),

unde log b = $\frac{\Sigma t_i \log y_i}{\Sigma t_i^2}$, pentru $\Sigma t_i = 0$, i = $\overline{1,n}$.

Ritmul mediu al sporului ($\bar{r}$). Ritmul mediu al sporului arată, în expresie relativă, cu cât sa modificat în medie pe an nivelul unui fenomen într-o perioadă.

Acest indicator se calculează după relația:

$$\bar{r} = \bar{R} - 1, sau \bar{r} = \bar{R} - 100 \qquad (9.18)$$

Exemplificare: Volumul comerțului exterior al României, în perioada 1981–1989, este prezentat în tabelul 9.1.

Scanned with OKEN Scanner

Tabelul 9.1. Volumul comerțului exterior al României,în perioada 1981–1989

Anii	Comerțul exterior total (în mil. lei valută)
1981	332373
1982	272572
1983	297929
1984	367417
1985	325054
1986	299770
1987	300834
1988	304521
1989	302762

Sursa: Anuarul statistic al României, C.N.S. 1990, p. 605

Calculul indicatorilor absoluți este redat în tabelul 9.2, coloanele 3 şi 4.

Tabelul 9.2. Valoarea sporurilor

Anii	Comerțul exterior (mil. lei valută) (y_i)	Sporul absolut (mil. lei valută)	
		Cu bază fixă $y_i - y_0$	Cu bază mobilă $y_i - y_{i-1}$
1	**2**	**3**	**4**
1981	332373	0	–
1982	272572	–59801	–59801
1983	297929	–34444	25357
1984	367417	35044	69488
1985	325054	–7319	–42363
1986	299770	–32603	–25284
1987	300834	–31539	1064
1988	304521	–27852	3687
1989	302762	–29611	–1759
TOTAL	**2803232**	–	**–29611**

Scanned with OKEN Scanner

Verificarea valorilor obținute se face pe baza relației dintre sporurile cu bază mobilă și cele cu bază fixă:

$$\Sigma \Delta_{i/i-1} = \Delta_{n/0}$$

Relația este verificată: suma valorilor din coloana 4 este egală cu valoarea de pe ultimul rând din coloana trei, tabelul 9.2.

Calculul indicatorilor relativi este redat în tabelul 9.3

Tabelul 9.3. Ritmurile variației și ale sporului

Anul	y_i (mil. lei)	Ritmul variației (%)		Ritmul sporului (%)	
		baza fixă	baza mobilă	baza fixă	Baza mobilă
		$(y_i/y_0)\cdot 100$	$(y_i/y_{i-1})\cdot 100$	$(\Delta_{i/0}/y_0)\cdot 100$	$(\Delta_{i/i-1}/y_{i-1})\cdot 100$
1981	332373	100,00	–	–	–
1982	272572	82,01	82,01	–17,99	–17,99
1983	297929	89,64	109,30	–10,36	9,30
1984	367417	110,54	123,32	10,54	23,32
1985	325054	97,80	88,47	–2,20	–11,53
1986	299770	90,19	92,22	–9,81	–7,78
1987	300834	90,51	100,35	–9,49	0,35
1988	304521	91,62	101,23	–8,38	1,23
1989	302762	91,09	99,42	–8,91	–0,58

Scanned with OKEN Scanner

Valoarea absolută a unui procent de creştere.

Valoarea absolută a unui procent de creştere cu bază fixă:

$$\%(y_{i/0}) = \frac{\Delta_{i/0}}{r_{i/0}}$$

$$\%(y_{i/0}) = \frac{\Delta_{1/0}}{r_{1/0}} = \frac{-59801}{-17,99} = 3324 \text{ mil. lei; } \%(y_{2/0}) =$$

$$= \frac{\Delta_{2/0}}{r_{2/0}} = \frac{-34444}{-10,36} = 3324 \text{ mil.lei}$$

$$\%(y_{10/0}) = \frac{\Delta_{10/0}}{r_{10/0}} = \frac{-29611}{-8,91} = 3324 \text{ mil.lei}$$

Valorile obținute arată că unui procent de creştere a comerțului exterior îi corespunde un spor absolut efal cu 3324 milioane lei.

Valoarea absolută a unui procent de creştere cu bază mobilă:

$$\%(y_{i/i-1}) = \frac{\Delta_{i/i-1}}{r_{i/i-1}}$$

$$\%(y_{1/0}) = \frac{\Delta_{1/0}}{r_{1/0}} = \frac{-59801}{-17,99} = 3324 \text{ mil. lei; } \%(y_{2/i}) =$$

$$= \frac{\Delta_{2/i}}{r_{2/i}} = \frac{-25357}{-9,3} = 2727 \text{ mil.lei}$$

$$\%(y_{8/7}) = \frac{\Delta_{8/7}}{r_{8/7}} = \frac{1759}{0,58} = 3024 \text{ mil. lei}$$

Calculul indicatorilor medii ai seriilor de timp

Nivelul mediu al seriei. Fiind o serie de timp de intervale, nivelul mediu se calculează ca o medie aritmetică, cu ajutorul relației:

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{2803232}{9} = 311470,22 \text{ mil. lei}$$

Scanned with OKEN Scanner

Sporul mediu ($\bar{\Delta}$):

$$\bar{\Delta} = \frac{\sum_{i=0}^{n} \Delta_{i/i-1}}{n} = \frac{\Delta_{n/0}}{n} = \frac{y_n - y_0}{n} = \frac{-29611}{8} = -3701{,}375 \text{ mil. lei}$$

Ritmul mediu al variației ($\bar{R}$):

$$\bar{R} = \sqrt[n]{\prod_{i-0}^{n} R_{i/i-1}} = \sqrt[n]{R_{n/0}} = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[8]{0{,}9109} = 0.9884 \text{ sau } 98{,}84\%$$

Ritmul mediu al sporului ($\bar{r}$):

$\bar{r} = \bar{R} - 1 = 0{,}9884 - 1 = -0{,}0116$ sau $\bar{r} = \bar{R} - 100 = 98{,}84 - 100 = -1{,}16\%$

Valoarea absolută a unui procent de creștere.

Valoarea absolută a unui procent de creștere cu bază fixă: $\%(y_{i/0}) = \frac{\Delta_{i/0}}{r_{i/0}}$

$$\%(y_{i/0}) = \frac{\Delta_{1/0}}{r_{1/0}} = \frac{-59801}{-17{,}99} = 3324 \text{ mil. lei;}$$

$$\%(y_{2/0}) = \frac{\Delta_{2/0}}{r_{2/0}} = \frac{-34444}{-10{,}36} = 3324 \text{ mil.lei}$$

$$\%(y_{10/0}) = \frac{\Delta_{10/0}}{r_{10/0}} = \frac{-29611}{-8{,}91} = 3324 \text{ mil.lei}$$

Valorile obținute arată că unui procent de creștere a comerțului exterior îi corespunde un spor absolut egal cu 3324 milioane lei.

Valoarea absolută a unui procent de creștere cu bază mobilă:

$$\%(y_{i/i-1}) = \frac{\Delta_{i/i-1}}{r_{i/i-1}}$$

$$\%(y_{1/0}) = \frac{\Delta_{1/0}}{r_{1/0}} = \frac{-59801}{-17{,}99} = 3324 \text{ mil. lei;}$$

Scanned with OKEN Scanner

$$\%(y_{2/i}) = \frac{\Delta_{2/i}}{r_{2/i}} = \frac{-25357}{-9,3} = 2727 \text{ mil.lei}$$

$$\%(y_{8/7}) = \frac{\Delta_{8/7}}{r_{8/7}} = \frac{1759}{0,58} = 3024 \text{ mil. lei}$$

Calculul indicatorilor medii ai seriilor de timp

Nivelul mediu al seriei. Fiind o serie de timp de intervale, nivelul mediu se calculează ca o medie aritmetică, cu ajutorul relației:

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{2803232}{9} = 311470,22 \text{ mil. lei}$$

Sporul mediu ($\bar{\Delta}$):

$$\bar{\Delta} = \frac{\sum_{i=0}^{n} \Delta_{i/i-1}}{n} = \frac{\Delta_{n/0}}{n} = \frac{y_n - y_0}{n} = \frac{-29611}{8} = -3701,375 \text{ mil. lei}$$

Ritmul mediu al variației ($\bar{R}$):

$$\bar{R} = \sqrt[n]{\prod_{i-0}^{n} R_{i/i-1}} = \sqrt[n]{R_{n/0}} = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[8]{0,9109} = 0.9884 \text{ sau } 98,84\%$$

Ritmul mediu al sporului ($\bar{r}$):

$\bar{r} = -1 = 0,9884 - 1 = -0,0116$ sau $\bar{r} = \bar{R} - 100 = 98,84 - 100 = -1,16\%$

9.2. **Indicii dinamicii**

Indicii statistici se utilizează pentru a caracteriza variația medie relativă a fenomenelor în timp, pentru a măsura influența diferiților factori asupra nivelului și structurii fenomenelor analizate.

Indicii au fost elaborați, inițial, cu scopul de a măsura variația prețurilor bunurilor materiale, astăzi se aplică în toate domeniile de activitate.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

9.2.1. **Categorii de indici**

Indicii statistici se calculează ca raport între două nivele ale unui fenomen, înregistrate fie pentru momente, fie pentru spaţii diferite.

După natura variaţiei pe care o măsoară, indicii statistici pot fi: indici ai dinamicii, indici teritoriali şi indici calitativi.

După baza de comparare, indicii pot fi calculaţi cu bază fixă sau cu bază mobilă.

După sfera de cuprindere a fenomenului format din mărimi simple sau din mărimi complexe (incluzând mărimi simple cu ponderi diferite), indicii statistici se grupează în două categorii: indici individuali şi indici de grup.

Indicii individuali, Indicii individuali, numiţi şi indici simpli, se calculează cu relaţiile:

$$i^{yi}_{1/0} = \frac{y_1}{y_0} \text{ sau } i^{yi}_{1/0} = \frac{y_1}{y_0} 100 \qquad (9.19)$$

unde: y_0 şi y_1 reprezintă nivelul fenomenului y înregistrat în perioada de bază, respectiv curentă;

$i^{yi}_{1/0}$ — indicele individual al fenomenului y

Indicii de grup. Sunt notaţi cu $i^{yi}_{1/0}$ şi arată variaţia medie în timp a elementelor unui ansamblu. Indicii de grup se pot calcula sub mai multe forme: indici agregaţi, indici calculaţi cu mărimi medii sau indici de grup calculaţi ca raport de medii.

9.2.2. **Indici agregaţi**

Indicii agregaţi se determină prin raportarea nivelului agregat al unui fenomen din două perioade comparate: de bază (0) şi curentă (1). Nivelul agregat al unui fenomen se obţine prin însumarea elementelor componente:

$\sum_i y_i$, pentru elemente omogene, respectiv $\sum_i y_i f_i$, pentru elemente eterogene.

Indicele agregat simplu şi indicele agregat ponderat se calculează după

$$i_{1/0}^{yi} = \frac{\sum_i y_{i1}}{\sum_i y_{i0}} \quad 100 \text{ şi } i_{1/0}^{y} = \frac{\sum_i y_{i1} f_{i1}}{\sum_i y_{i0} f_{i0}} \cdot \quad 100 \qquad (9.20)$$

9.2.3. **Indici calculaţi ca mărimi medii**

Indicii calculaţi ca mărimi medii se aplică atunci când se cunosc valorile indicilor individuali şi ponderea complexă (y f) numai dintr-o perioadă.

Indicele mediu aritmetic: $$i_{1/0}^{y} = \frac{\sum_i i_{1/0}^{y} \cdot y_{i0} f_{i0}}{\sum_i y_{i0} f_{i0}} \qquad (9.21)$$

Indicele mediu armonic: $$i_{1/0}^{y} = \frac{\sum_i y_{i1} f_{i1}}{\sum_i i_{1/0} \frac{1}{y} \cdot y_{i1} f_{i1}} \qquad (9.22)$$

Indicele mediu geometric – Irving Fischer:

$$i_{1/0}^{y} = \sqrt{\frac{\sum_i y_{i1} f_{i0}}{\sum_i y_{i0} f_{i0}} \cdot \frac{\sum_i y_{i1} f_{i1}}{\sum_i y_{i0} f_{i1}}} \qquad (9.23)$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

9.2.4. **Indici calculaţi ca raport de medii**

Indicii calculaţi ca raport de medii se mi numesc şi **indici ai valorilor medii** şi prezintă dinamica nivelului mediu al unui fenomen, precum şi dinamica elementelor componente ale valorii medii.

Indicii din această categorie se calculează după relaţia:

$$I_{1/0}^{\bar{y}} = \frac{\overline{y}_1}{\overline{y}_0} \cdot 100 \qquad (9.24)$$

În funcţie de elementele componente ale valorii medii a fenomenului studiat se calculează trei tipuri de indici cu roluri analitice diferite.

Indicele bifactorial: $$I_{1/0}^{(y,f)} = \frac{\sum_i y_{i1} f_{i1}}{\sum_i f_{i1}} : \frac{\sum_i y_{i0} f_{i0}}{\sum_i f_{i0}} \qquad (9.25)$$

Indicele cu structură fixă:

$$I_{1/0}^{y} = \frac{\sum_i y_{i1} f_{i1}}{\sum_i f_{i1}} : \frac{\sum_i y_{i0} f_{i1}}{\sum_i f_{i1}} = \frac{\sum_i y_{i1} f_{i1}}{\sum_i y_{i0} f_{i1}} \qquad (9.26)$$

Indicele cu structură variabilă:

$$I_{1/0}^{(f)} = \frac{\sum_i y_{i0} f_{i1}}{\sum_i f_{i1}} : \frac{\sum_i y_{i0} f_{i0}}{\sum_i f_{i0}} \qquad (9.27)$$

Dacă în indicii prezentaţi se face substituţia $G = \frac{f}{\Sigma f}$, se obţine forma agregată a acestora:

$$I_{1/0}^{(y,f)} = \frac{\sum_i y_{i1} G_{i1}}{\sum_i y_{i0} G_{i0}}; I_{1/0}^{(y)} \frac{\sum_i y_{i1} G_{i1}}{\sum_i y_{i0} G_{i1}}; I_{1/0}^{(f)} = \frac{\sum_i y_{i0} G_{i1}}{\sum_i y_{i0} G_{i0}} \qquad (9.28)$$

Relații între indicii calculați ca raport de medii:

$$I_{1/0}^{(y,f)} = I_{1/0}^{(y)} \cdot I_{1/0}^{(f)} \qquad (9.29)$$

de unde se poate afla $I_{1/0}^{y}$ sau $I_{1/0}^{f}$ când se cunosc ceilalți doi.

9.2.5. **Indici după felul bazei de calcul**

În această categorie sunt incluşi indicii cu bază mobilă sau în lanț.

Pentru indicii cu bază fixă, alegerea bazei de calcul trebuie să țină seama de schimbările din viața economică (schimbări de structuri, evoluții instituționale, progres tehnic) pentru a se reflecta corespunzător evoluția fenomenului în timp. Indicii după felul bazei de raportare sunt tratați ca ritmuri ale variației, cu bază fixă şi cu bază mobilă, şi anume:

$$I_{1/0}^{y} = \frac{y_i}{y_0}; \; I_{1/i-1}^{y} = \frac{y_i}{y_{i-1}} \qquad (9.30)$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 10

ANALIZA ŞI EXTRAPOLAREA SERIILOR CRONOLOGICE

Studiul seriilor cronologice presupune *analiza dinamicii unei variabile statistice* în scopul descrierii, modelării şi extrapolării acesteia în timp pe componente definitorii. Această metodă pleacă de la analiza trecutului şi vizează extrapolarea tendinţelor manifestate în perioada observată, bazându-se pe următoarele ipoteze:

— tendinţele observate în trecut se vor menţine în viitor;

— fluctuaţiile unei variabile se reproduc la intervale regulate.

Simultan cu interesul manifestat pentru dinamica fenomenelor s-au dezvoltat noi tehnici de analiză a seriilor cronologice, de la vizualizarea grafică, la analiza armonică, analiza de regresie, analiza fractală. O contribuţie deosebită, în acest domeniu, a adus W.M. Pearson (1919), care a descompus o serie cronologică în componente — tendenţială, ciclică, sezonieră şi accidentală. Apoi, numeroase lucrări au fost consacrate determinării şi eliminării uneia sau alteia din componente.

10.1. Tipologia seriilor cronologice

10.1.1. Componentele unei serii cronologice

Orice serie cronologică poate fi descompusă în patru componente, fiecare exprimând un aspect particular al valorilor seriei, definind şi tipologia lor. Aceste componente sunt: *tendinţa, fluctuaţia ciclică, variaţia sezonieră şi variaţia aleatoare, prezentate grafic în figura 10.1.*

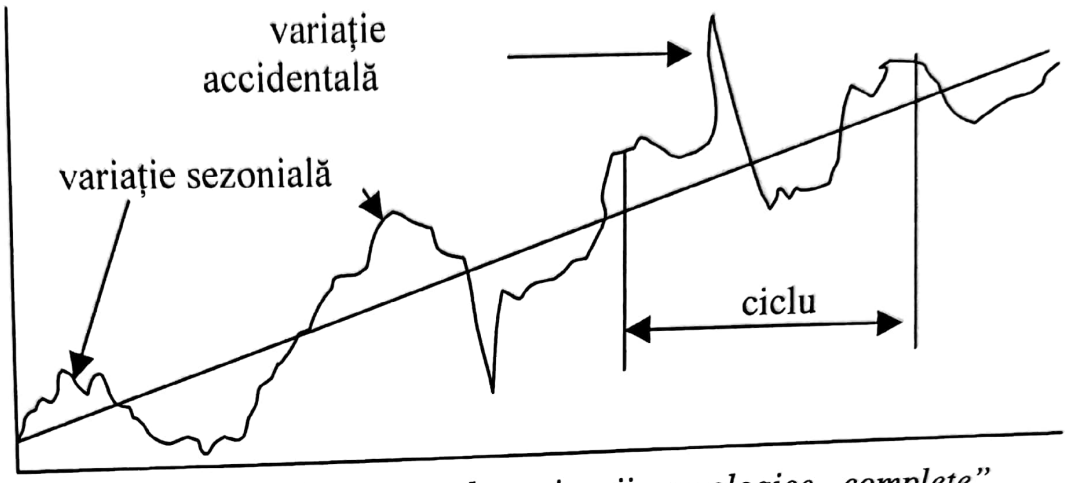


Figura 10.1. Componentele unei serii cronologice „complete"

10.1.2. **Componenta tendențială (trend)**

Componenta tendențială, numită şi *trend* sau *tendință seculară*, se notează f_t şi *arată tendința generală înregistrată pe o perioadă lungă de timp şi exprimă variația medie*, respectiv *legea de evoluție* a variabilei observate. Este de natură deterministă, timpul fiind singura variabilă care explică variația fenomenului observat.

Descrierea componentei trend se face printr-o *funcție continuă*, de regulă, printr-o funcție polinomială în timp, $y_t = f(t)$, de grad n(cu n=1 sau 2, foarte rar n>2). Aproximarea modelului de trend (a legii de evoluție în timp a variabilei observate) şi estimarea parametrilor acestuia se constituie ca obiect al ajustării statistice.

10.1.3. **Componenta ciclică**

Componenta ciclică se notează C_t şi *se prezintă ca o fluctuație de tip sinusoidal, în jurul trendului*, cu o durată relativ lungă de timp, cu caracter de repetabilitate regulată. Variația ciclică cuprinde patru faze: expansiunea, criza, recesiunea şi relansarea (vezi figura 10.2.)

Scanned with OKEN Scanner

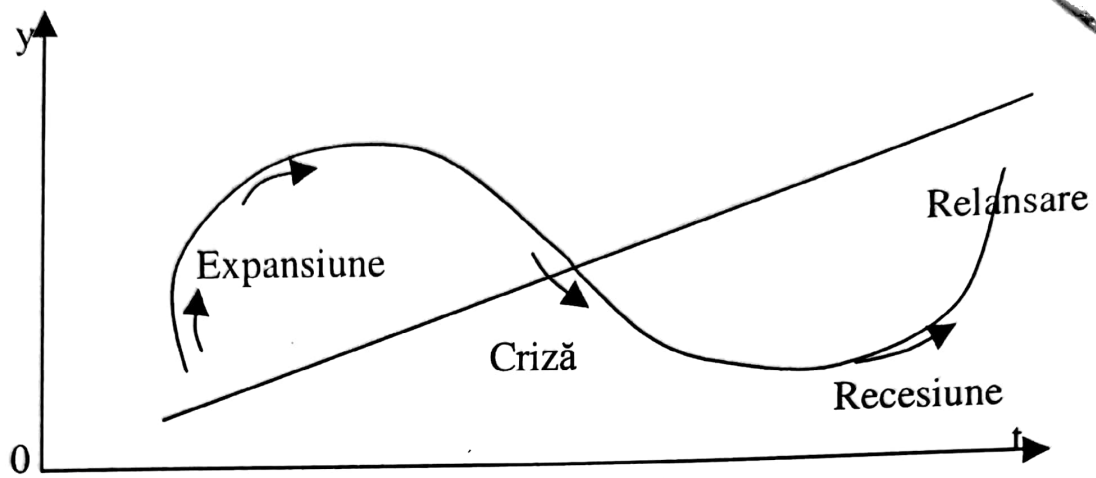


Figura 10.2. Fazele unui ciclu

În activitatea economică s-au conturat cicluri cu durată diferită, de aproximativ 50 ani pentru ciclul tip Kondratieff, de aproximativ 9 ani pentru ciclul tip Juglar, de aproximativ 7 ani pentru ciclul biblic.

Analiza componentei ciclice (atunci când aceasta există, deoarece adesea se suprapune pe trend) presupune studierea comportamentului variabilei observate perioadă lungă de timp în raport cu trendul, decupându-se periodicitatea ciclică şi fazele unui ciclu.

10.1.4. **Componenta sezonieră**

Componenta sezonieră se notează S_t este definită de variaţiile pe termen scurt, prin creşteri şi descreşteri de nivel succesive care se repetă de la o perioadă la alta, fie în jurul componentei ciclice (când aceasta există) fie în jurul componentei trend.

Variaţiile sezoniere sunt repetabile, de regulă, de la o lună la alta, sau de la un trimestru la altul. La nivelul unui an influenţa sezonieră este neutră.

Variaţiile sezoniere, S_t , se repetă riguros identic de la o perioadă „p" la alta, adică:

$S_t = S_{t+2p} = S_{t+2p} = \ldots$ = etc.

Scanned with OKEN Scanner

Dacă, de exemplu, perioadele ar fi lunile sau trimestrele anului, atunci variațiile sezoniere corespunzătoare se repetă identic lunar sau trimestrial în fiecare an, respectiv, au loc egalitățile:

$S_t = S_{t+12p}$ adică $S_1 = S_{25} = \ldots =$ etc. sau

$S_t = S_{t+4}$, adică $S_1 = S_5 = S_9 = \ldots$

Influența variațiilor sezoniere S_t (lunare sau trimestriale) *este neutră la nivelul anului*; creşterile şi descreşterile de nivel (lunare sau trimestriale) se compensează între ele la nivelul fiecărui an. Ca urmare, au loc următoarele relații:

— media variațiilor sezoniere este nulă la nivelul anului:

$$\sum_{t=1}^{p} S_i = 0, \text{ respectiv, } \overline{S}_t = 0, \overline{S} = \frac{1}{p} = \sum_{t=1}^{p} S_t, \qquad (10.1)$$

În cazul unui model aditiv, (considerând $\Sigma e_t = 0$), respectiv

— media variațiilor sezoniere este egală cu unitatea:

$$\overline{S} = 1, \overline{S} = \frac{1}{p} \sum_{t=1}^{p} S_t \qquad (10.2)$$

în cazul unui model multiplicativ, (considerând $\Sigma e_t = 0$)

În cazul modelului multiplicativ, variațiile sezoniere cresc sau descresc cu aceeaşi rată (proporție), s, astfel că:

unde $1+\overline{s}$ este *multiplicatorul*, iar $\overline{s}$ rata medie de creştere, egală cu zero: $\overline{s} = 0$. (10.3)

Adică, analog modelului aditiv, rata de creştere a variațiilor sezoniere se anulează, în medie, în fiecare an, repetându-se de la an la an.

10.1.5. **Componenta aleatoare (reziduală)**

Componenta aleatoare se notează e_t şi are drept caracteristică de bază caracterul non-determinist al variației. Se consideră că pe un număr mic de ani e_t se compensează $\Sigma e_t = 0$. Variația aleatoare se poate manifesta ca *un proces pur aleator, un proces aleator în care parametrii variază în timp şi ca un proces staționar.*

Scanned with OKEN Scanner

Scanned with OKEN Scanner

În cazul procesului pur aleator seria cronologică este o secvență de variabile independente mutual şi cu aceeaşi distribuție de probabilitate. Speranța matematică şi varianța variabilei aleatoare e_t sunt independente de timp:

$E(e_t) = \mu$ şi $V(e_t) = \sigma_e^2$, oricare ar fi perioada t, iar covarianța a două perturbații aleatoare decalate de un număr j de perioade este nulă:

$Cov\ (e_t, e_{t+j}) = 0$, oricare ar fi j.

Un proces pur aleator mai este numit **zgomot alb**, denumire ce implică necesar normalitatea perturbațiilor aleatoare (sub această ipoteză s-au determinat estimatorii corecți ai parametrilor necunoscuți ai unei populații, precum şi estimatorii parametrilor de regresie).

În cazul proceselor aleatoare în care parametrii variază în timp se includ procese aleatoare ale căror realizări sunt independente de la o perioadă la alta dar ai căror parametri evoluează în cursul timpului. Fiecare dată a seriei este o realizare a unei variabile diferită de precedentele. Respectă aceleaşi proprietăți specifice proceselor aleatoare:

$$E(e_t) = \mu;\ V(e_t) = \sigma_{e_t}^2,\ cov\ (e_t, e_{t+j}) = 0 \qquad (10.4)$$

În cazul proceselor staționare caracteristică este dependența variabilei de e_{t-1}, $e_t-2,\ldots$, e_{t-k}, dependență exprimată prin proprietatea: covarianța între e_t şi e_{t-1} rămâne constantă şi nu depinde decât de decalajul j şi nu de t. Ca urmare proprietățile specifice proceselor staționare, în raport cu procesele pur aleatoare, sunt,

$$E(e_t) = \mu;\ V(e_t) = \sigma_e^2, \text{ pentru oricare t,} \qquad (10.5)$$

$cov\ (e_t, e_{t-j}) \neq 0$, rămâne constantă şi nu depinde decât de j şi nu de t.

Scanned with OKEN Scanner

10.2. **Ajustarea seriilor cronologice**

Ajustarea seriilor de timp presupune înlocuirea termenilor seriei date cu termeni teoretici care exprimă tendința de dezvoltare a unui fenomen.

Tendința (trend-ul sau linia medie de dezvoltare) reflectă legitatea specifică de dezvoltare a unui fenomen pe o perioadă lungă de timp, făcând abstracție de abaterile față de nivelul mediu, respectiv, de eroarea sau reziduul datorate influenței factorilor aleatori.

10.2.1. **Metoda grafică de ajustare**

Ajustarea prin metoda grafică este preliminară oricărei alte metode de ajustare. Pin această metodă se aproximează vizual modelul de evoluție care corespunde cel mai bine tendinței de dezvoltare a fenomenului.

Aplicarea acestei metode *constă în construirea cronogramei şi trasarea vizuală a liniei de tendință*. Aproximarea grafică a trendului presupune trasarea unei linii continue, în sensul de variație a variabilei timp, care să treacă prin sau prin apropierea a cel puțin două puncte de coordonate (t_i, y_i), $i = \overline{1,n}$ şi să împartă norul de puncte în două părți aproximativ egale, jumătate din puncte deasupra liniei de trend, jumătate sub linie (fig. 10.3.).

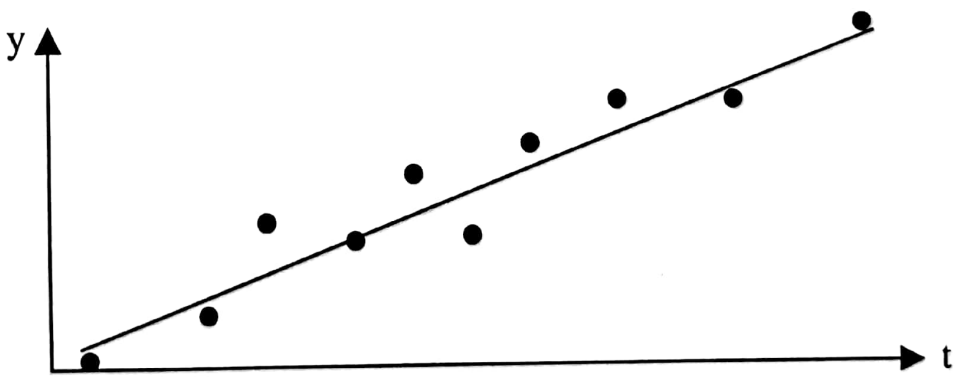


Figura 10.3. Ajustarea grafică

Scanned with OKEN Scanner

10.2.2. **Metoda analitică de ajustare**

Metoda analitică de ajustare *constă în aproximarea tendinței de variație în timp cu ajutorul funcției de trend* $y_t = f(t_i)$, în care "t" reprezintă variabila timp, iar „y" variabila în timp. Pin folosirea metodei analitice de ajustare se stabileşte cu mai multă siguranță legea de dezvoltare, de variație a unui fenomen, deoarece se ține seama de valorile empirice ale tuturor termenilor unei serii.

Tendința de variație a fenomenelor sociale se aproximează, de regulă, printr-una din funcțiile ale căror curbe şi ecuații de estimare au fost prezentate în capitolul „Regresie şi corelație". Variabila x din modelele de regresie devine t în modelele de trend.

Parametrii ecuației de trend se află, de regulă, cu ajutorul metodei celor mai mici pătrate.

Trendul liniar. O variație în timp de forma unei funcţii liniare este specifică fenomenelor care se dezvoltă rectiliniu. Ecuația de estimare a funcției de trend liniare este definită de relația:

$$Y_t = a+bt_i+e \qquad (10.6.)$$

Determinarea parametrilor funcției de trend are la bază principiul celor mai mici pătrate, conform căruia trebuie să se satisfacă condiția:

$$S = \sum e_i^2 = \sum(y_i - \bar{y}_t)^2 = \text{minim} \qquad (10.7)$$

Pentru un trend liniar, principiul celor mai mici pătrate cere să se satisfacă condiția:

$$S = \sum(y_i - a - bt_i)^2 = \text{minim} \qquad (10.8)$$

Se stabilesc şi se anulează derivatele parţiale ale expresiei „S" în raport cu a şi b:

$$\frac{\partial S}{\partial S} = 2\sum(y_i - a - b\,t_i)\,(-1) = 0 \qquad (10.9)$$

$$\frac{\partial S}{\partial b} = 2\sum(y_i - a - b\,t_i)\,(-t_i) = 0 \qquad (10.10)$$

De aici rezultă următorul sistem de două ecuații normale:

$$n \cdot a + b\sum t_i = \sum y_i \qquad (10.11)$$
$$a \cdot \sum t_i + b\sum t_i^2 = \sum t_i y_i$$

În condițiile funcției de trend, variabila timp reprezintă de regulă o serie de numere consecutive, originea luându-se în centrul seriei şi deci $\sum t_i = 0$. Ca urmare, ecuațiile normale pot fi simplificate, luând următoarea formă:

$$n \cdot a = \sum y_i \qquad (10.12)$$
$$b\sum t_i^2 = \sum t_i y_i$$

de unde rezultă relațiile de calcul pentru parametrii ecuației de trend:

$$a = \frac{\Sigma y_i}{n} \text{ şi } b = \frac{\Sigma t_i y_i}{\Sigma t_i^2} \qquad (10.13)$$

Valorile variabilei timp se măsoară în unități întregi (intervale), dacă numărul termenilor seriei este impar şi anume:

..., –2, –1, 0, 1,2, ...

sau în jumătăți de interval, dacă numărul termenilor este par, astfel:

..., –5, –3, –1, 0, 1, 3, 5,...

Aflarea parametrilor se efectuează mai facil dacă elementele de calcul sunt grupate într-un tabel (vezi tabelul 10.1.)

Exemplificare. Considerând datele prezentate în tabelul 10.1. cu un fenomen y, într-o perioadă de 11 ani, se cere să se ajusteze seria prin metoda analitică.

Scanned with OKEN Scanner

Tabelul 10.1. Elemente de calcul pentru determinarea tendinței liniare

Anii	y_i	t_i	$t_i y_i$		Valorile seriei ajustate (teoretice) $y_t = a + bt_i = 56,9 + 9,44t_i$
1	2	3	4	5	6
1	22,9	–5	–114,5	25	14,7
2	25,2	–4	–180,8	16	23,14
3	28,8	–3	–86,4	9	31,58
4	36,0	–2	–72,	4	40,82
5	49,8	–1	–49,8	1	48,46
6	53,1	0	0	0	56,9
7	60,8	1	68,8	1	65,34
8	69,6	2	139,2	4	73,78
9	77,4	3	232,2	9	82,22
10	92,3	4	369,2	16	90,66
11	1	5	550,8	25	99,10
	10,0				
Total	**625,9**	**0**	**927,9**	**110**	**625,9**

Introducând datele, calculate în tabelul 10.1., în sistemul (10.12) se obține:

11a = 625,9

110b = 927,9

de unde rezultă valoarea parametrilor:

a = 56,9 şi b = 8,44Ecuaţia tendinţei liniare va fi:

$Y_t = a + bt_i = 56,9 + 8,44\ t_i$,

iar dreapta determinată pe baza acestei ecuaţii este prezentată în coloana 6 a tabelului 10.1.

Verificarea exactității ajustării constă în compararea sumei valorilor empirice cu suma valorilor ajustate (teoretice) ale termenilor seriei, care trebuie să fie egale ($\Sigma y_i = \Sigma y_t$), în ipoteza $\Sigma e_i = 0$.

Scanned with OKEN Scanner

În seria considerată, abaterile în plus şi în minus faţă de tendinţa medie sunt relativ mici, ceea ce ne îndreptăţeşte să credem că dreapta calculată oglindeşte corect trendul în perioada cercetată.

Trendul parabolic. Trendul parabolic este specific fenomenelor care prezintă o tendinţă crescătoare sau descrescătoare cu un punct de maxim, respectiv de minim.

Ecuaţia tendinţei parabolice de gradul doi este:

$$n \cdot a + b \cdot \Sigma t_i + c \cdot \Sigma t_i^2 = \Sigma y_i$$

$$a \cdot \Sigma t_i + b \cdot \Sigma t_i^2 + c \cdot \Sigma t_i^3 = \Sigma t_i y_i$$

$$a \cdot \Sigma t_i^2 + b \cdot \Sigma t_i^3 + c \cdot \Sigma t_i^4 = \Sigma t_i^2 y_i$$

în condiţia $\Sigma t_i = 0$, se obţin:

$$a = \frac{\Sigma y_i \cdot \Sigma t_i^4 - \Sigma t_i^2 y_i \cdot \Sigma t_i^2}{n \cdot \Sigma t_i^4 - (\Sigma t_i^2)^2} \qquad (10.15)$$

$$b = \frac{\Sigma t_i y_i}{\Sigma t_i^2} \qquad (10.16)$$

$$c = \frac{n \cdot \Sigma t_i^2 y_i / \Sigma t_i^2 \cdot \Sigma y_i}{n \cdot \Sigma t_i^4 - (\Sigma t_i^2)^2} \qquad (10.17)$$

Considerăm datele din tabelul 10.1. pentru ajustare după un trend parabolic.

Elementele de calcul sunt prezentate în tabelul 10.2. Înlocuind datele din tabelul 10.2. în sistemul de ecuaţii, se obţine soluţia sistemului:

$11a + 110c = 625,9$ $\quad b = 8,435454 \approx 8,44$

$110b = 927,9$ $\quad a = 51,861538 \approx 51,86$

$110a + 1958c = 6691,3$ $\quad c = 0,503846 \approx 0,504$

Scanned with OKEN Scanner

Tabelul 10.2. Elemente de calcul

Anii	y_i	t_i	t_i^2	t_i^4	$t_i y_i$	$t_i^2 y_i$	Seria ajustată $Y_t = a + bt_i + ct_i^2 = 51,86 + 8,44t_i + 0,504t_i^2$
1	2	3	4	5	6	7	6
0	22,9	–5	25	625	–114,5	572,5	22,26
1	25,2	–4	16	256	–100,8	403,2	26,164
2	28,8	–3	9	81	–86,4	259,2	31,076
3	36,0	–2	4	16	–72,0	144,0	36,996
4	49,8	–1	1	1	–49,8	49,8	43,924
5	53,1	0	0	0	0	0	51,06
6	60,8	1	1	1	60,8	60,8	68,804
7	69,6	2	4	16	139,2	278,4	70,756
8	77,4	3	9	81	232,2	696,6	81,716
9	92,3	4	16	256	369,2	1476,8	93,684
10	110,0	5	25	625	550,0	2750,0	106,66
11							
	625,9	**0**	**110**	**1958**	**927,9**	**6691,3**	**625,906**

Ecuația tendinței parabolice pentru seria dată va fi:

$$y_t = a + bt_i + ct_1^2 = 51,86 + 8,44\,t_1 + 0,504\,t_1^2,$$

iar valorile teoretice pentru perioada I-II sunt prezentate în coloana a 8-a din tabelul 10.2.

Trendul exponențial. Trendul exponențial este specific fenomenelor care se dezvoltă asemănător unei progresii geometrice.

Ecuația de estimare a trendului exponențial este dată de relația:

$$y_t = a \cdot b^{ti} \qquad (10.18)$$

Prin logaritmare expresiei funcției de trend se obține:

$$\log y_t = \log a + t_i \log b \qquad (10.19)$$

Scanned with OKEN Scanner

Aplicând metoda celor mai mici pătrate în ajustarea după o curbă exponențială, a cărei ecuație a fost logaritmată și considerând condiția $\Sigma t_i = 0$,

se obțin formulele de calcul ale parametrilor modelului de trend:

$$\log a = \frac{\Sigma \log y_i}{n} \text{ și } \log b = \frac{\Sigma t_i \log y_i}{\Sigma t_i^2} \qquad (10.20)$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 11

ANALIZA DISPERSIONALĂ

Analiza dispersională[1] modelează studiile experimentale complexe, în care intervin colectivităţi formate din mai multe eşantioane sau grupe, fiecare dintre acestea ilustrând adesea mai multe situaţii sau condiţii experimentale.

În cazul general, vom considera o colectivitate formată, după diverse criterii, din *s* grupe distincte diferenţiate prin intervenţia unuia sau mai multor factori experimentali (ANOVA unifactorială, bifactorială, multifactorială). Spre deosebire de analiza dispersională, în cadrul inferenţei statistice „clasice" testele de semnificaţie operează cu cel mult două eşantioane.

În toate situaţiile, analiza dispersională operează cu mulţimi sau colectivităţi de grupe *independente* sau *dependente (perechi)*. Fiecare grup în sine constituie un *eşantion aleator*, format din subiecţi extraşi la întâmplare dintr-o populaţie statistică originară.

Grupele sau eşantioanele ca atare pot fi luate la întâmplare sau, adesea, pot fi alese pe baza unui criteriu bine determinat. În acest fel, asupra fiecărei grupe se exercită o influenţă determinată, pe baza unui plan sistematic de experimentare. Studiul urmăreşte în acest caz diferenţele dintre eşantioane rezultate în urma condiţiilor specifice diferite aplicate fiecăruia. Diferenţele specifice între eşantioane pot fi generate de influenţa unuia sau mai multor factori de diferenţiere.

---

[1] (*) Analiza dispersională este numită şi analiză de varianţă. Adeseori se foloseşte termenul ANOVA, acronimul denumirii în limba engleză („analysis of variance").

11.1 Noţiuni generale

Analiza dispersională analizează variaţia datelor (împrăştierea lor în jurul mediei). Metoda se bazează pe cunoaşterea surselor de variabilitate a valorilor obţinute în urma experimentului şi pe posibilitatea măsurării precise a rezultatelor (se aplică variabilelor măsurate la nivelul de interval-raport).

Colectivitatea experimentala este formata din grupe sau eşantioane *independente*. Cazul eşantioanelor dependente — perechi va fi analizat ulterior in paragraful 11.4.

Exemplu:

Se recurge la un experiment pentru a determina efectul unei investiţii asupra venitului net, bazat pe testarea a 4 grupuri de firme(un grup de control şi 3 grupuri experimentale care primesc investiţii diferite). Rezultatele înregistrate sunt sintetizate în tabelul următor:

	Gr. Experimental 1	Gr. experimental 2	Gr. experimental 3	Gr. control
Venituri nete înregistrate	69, 72, 81, 75, 70, 76	76, 81, 84, 80, 82, 79	80, 88, 91, 79, 85	68, 70, 67, 73, 71
Media eşantionului	$\overline{X}_1$=73,83	$\overline{X}_2$=80,33	$\overline{X}_3$=84,60	$\overline{X}_4$=69,80

Se observă că investiţia are influenţă asupra venitului net al grupurilor deoarece rezultatele celor 4 grupuri sunt în medie diferite, iar grupul de control (unde nu s-a administrat investiţia) are venitul net cel mai mic.

În această problemă pot fi formulate două ipoteze statistice:

— **ipoteza de nul**, care atribuie diferenţele dintre eşantioane întâmplării (mărimea diferenţelor este nesemnificativă);

Scanned with OKEN Scanner

— **ipoteza specifică**, diferențele dintre eşantioane sunt considerate semnificative, fiind produse ca urmare a valorilor diferite ale variabilei de diferențiere (mărimea investiției) pentru fiecare grup.

Pentru testarea acestor ipoteze se apelează la metodele analizei dispersionale. Notăm că asupra grupelor intervine un singur factor de diferențiere (mărimea investiției), deci exemplul prezentat este o problemă ANOVA unifactorială.

Evident, nu toți subiecții incluşi într-un experiment obțin acelaşi scor, rezultatele lor fiind dispersate în jurul mediei colectivității studiate. Această variabilitate (dispersie) a rezultatelor se înregistrează, în general, datorită intervenției simultane a două surse de variație:

- **Variabilitatea internă a grupelor**, datorată diversității măsurărilor şi a firmelor incluse (aleator) în fiecare grupă. Aceasta se va numi *variabilitatea intragrupe* şi reprezintă dispersia datelor în cadrul fiecărui eşantion față de media acestuia. Deoarece studiul urmăreşte diferențele specifice dintre eşantioane (apărute ca urmare a intervenției factorilor sistematici de diferențiere) şi nu interesează diversitatea subiecților incluşi în fiecare dintre grupuri, această variabilitate este o sursă de eroare care afectează rezultatele şi este numită şi *variabilitate eroare.*

Pentru fiecare din cele *k* eşantioane, variabilitatea eroare se scrie:

$$V_E^{(k)} = \sum_{i=1}^{N_k} \left(X_i^{(k)} - \overline{X}^{(k)} \right) \quad , k = 1,2,3... \qquad (11.1)$$

unde $X_i^{(k)}$ reprezintă valorile din eşantionul *k*, $\overline{X}^{(k)}$ este valoarea mediei aritmetice a eşantionului *k*, iar N_k volumul acestuia.

În cadrul întregii colectivități, variabilitatea intragrupe se defineşte ca suma variabilităților interne ale fiecărui eşantion din care se compune colectivitatea de studiu:

$$V_E = \sum_{k=1}^{s} V_E^{(k)} \ , \qquad (11.2)$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

s reprezentând numărul de eşantioane din care este alcătuită colectivitatea.

- **Variaţia dintre grupe**, reprezentată de „dispersia" mediilor eşantioanelor faţă de media generală a întregii colectivităţi (*variabilitatea intergrupe*). Această sursă de variabilitate apare ca urmare a condiţiilor diferite de experimentare pentru fiecare eşantion care conduc la rezultate, în medie, diferite între grupe.

  Cunoscând că media generală a celor *s* eşantioane este

$$\overline{\overline{X}} = \frac{\sum_{k=1}^{s} \left(N_k \cdot \overline{X}^{(k)} \right)}{\sum_{k=1}^{s} N_k}, \qquad (11.3)$$

variabilitatea intergrupe generată de un factor de diferenţiere *A* se defineşte prin formula:

$$V_A = \sum_{k=1}^{s} \left[\left(\overline{X}^{(k)} - \overline{\overline{X}} \right) \cdot N_k \right] \qquad (11.4)$$

Deoarece în general diferenţele specifice între eşantioane apar ca urmare a intervenţiei mai multor factori de diferenţiere (notaţi generic *A,B,...*), variabilitatea intergrupe rezultă din suma variaţiilor induse de fiecare dintre aceştia:

$$V_I = V_A + V_B + \ldots \qquad (115)$$

- **Variabilitatea totală** a datelor este definită prin împrăştierea valorilor din întreaga colectivitate (din toate cele *s* eşantioane) în jurul mediei acesteia (media generală):

$$V_{tot} = \sum_{k=1}^{s} \left(\sum_{i=1}^{N_k} \left(X_i^{(k)} - \overline{\overline{X}} \right) \right) \qquad (11.6)$$

După cum am văzut, această dispersie a datelor provine din două surse distincte (variabilitatea internă a grupelor şi variabilitatea intergrupe). Cu precizările anterioare, variabilitatea totală se poate scrie:

$$V_{tot} = V_E + V_I \qquad (11.7)$$

Pentru testarea ipotezelor statistice (şi extinderea aspectelor relevate de studiul colectivităţii testate la populaţia originară) în problemele de analiză dispersională se apelează la **testul Fischer — Snedecor** bilateral. Această procedură presupune distribuţia normală a rezultatelor în populaţia generală din care s-au extras eşantioanele, precum şi omogenitatea dispersiilor.

11.2 Analiza dispersională unifactorială

În acest tip de probleme experimentul are ca scop să desprindă influenţa unui singur factor de diferenţiere (sau clasificare a grupelor) prin *efectele* sau variaţiile produse de acesta în ansamblul rezultatelor înregistrate. Colectivitatea de studiu se constituie din s grupe independente după regulile selecţiei aleatoare.

Să analizăm experimentul din exemplul anterior. Colectivitatea experimentală cuprinde 4 grupuri (un grup de control şi 3 grupuri experimentale care primesc diferite investiţii), diferenţiate prin intervenţia unei singure variabile — mărimea investiţiei.

	Gr. experimental 1	Gr. experimental 2	Gr. experimental 3	Gr. control
Venituri nete înregistrate	69, 72, 81, 75, 70, 76	76, 81, 84, 80, 82, 79	80, 88, 91, 79, 85	68, 70, 67, 73, 71

În această problemă pot fi formulate două ipoteze statistice:

— **ipoteza de nul** (H_0): diferenţele dintre eşantioane sunt nesemnificative şi pot fi atribuite întâmplării. Grupele sunt omogene, provin din aceeaşi populaţie statistică sub aspectul venitului net şi efectul mărimii investiţiei este nesemnificativ, deci $m_1 = m_2 = m_3 = m_4$).

— **ipoteza specifică** (H_1): diferenţele dintre eşantioane sunt semnificative, cel puţin două eşantioane provenind din populaţii statistice diferite

Scanned with OKEN Scanner

în privința variabilei studiate. La nivelul populației mărimea investiției are un efect semnificativ asupra venitului net, deci $\exists\ m_i \neq m_j$.

Exemplul se referă la o colectivitate experimentală formată din N=22 subiecți grupați în $s = 4$ eşantioane.

Pentru testarea ipotezelor precizate se calculează raportul F (Fischer — Snedecor):

$$F_{calc} = \frac{MS_I}{MS_E} \qquad (11.8)$$

Acest raport are o distribuție statistică bine precizată, caracterizată de două „grade de libertate". Valoarea calculată se compară cu valoarea critică obținută din tabelul distribuției F pentru gradele de libertate $l_1 = s - 1$ (valoare corespunzătoare estimatorului MS_I) şi $l_2 = N - s$ (corespunzătoare MS_E) la pragul de semnificație impus.

Conform regulilor de decizie pentru testele bilaterale, în cazul $F_{calc} > Fcr$ se respinge ipoteza de nul (H_0) şi se acceptă ipoteza specifică (H_1). Cazul $F_{calc} < F_{cr}$ conduce la invalidarea testului prin starea de indecizie rezultată (nu se respinge ipoteza nulă, deci nu se poate accepta ipoteza specifică).

Valorile MS_I si MS_E care intervin in formula (11.8) se numesc *estimatori* şi se calculează prin raportarea variabilităților intergrupe, respectiv eroare (precizate anterior) la gradele de libertate corespunzătoare:

$$MS_I = \frac{V_I}{s-1}, \qquad MS_E = \frac{V_E}{N-s} \qquad (11.9)$$

unde variabilitățile V_I si V_E se calculează cu formulele (11.2) respectiv (11.4). Justețea calculelor se poate verifica cu formula (11.7), conform principiului prezentat anterior — variabilitatea totală este suma variabilităților intragrupe şi intergrupe.

Scanned with OKEN Scanner

În cele ce urmează va fi prezentată o alta metodă de calcul a variabilităţilor, care necesită un efort computaţional redus şi se pretează foarte bine la rezolvarea cu ajutorul unor programe de calcul sau cu modulul de calcul statistic al calculatoarelor. Setul de formule este întrutotul echivalent cu exprimările anterioare, putând fi dedus din acestea cu ajutorul unor transformări matematice relativ simple.

Pentru început vor fi calculate suma (totalul) valorilor, respectiv suma pătratelor valorilor din fiecare eşantion:

Totalul valorilor din eşantionul *k*:

$$T^{(k)} = \sum_{i=1}^{N_k} X_i^{(k)} \qquad (11.10)$$

Suma pătratelor valorilor din eşantionul *k*:

$$SP^{(k)} = \sum_{i=1}^{N_k} \left(X_i^{(k)}\right)^2 \qquad (11.11)$$

Variabilitatea eroare se calculează cu formula:

$$V_E = \sum_{k=1}^{s} SP^{(k)} - \sum_{k=1}^{s} \frac{\left(T^{(k)}\right)^2}{N_k} \equiv SP - R \ , \qquad (11.12)$$

unde *R* este suma rapoartelor de tipul $R^{(k)} = \left(T^{(k)}\right)^2 / N_k$:

$$R = \sum_{k=1}^{s} \left[\left(T^{(k)}\right)^2 / N_k\right] \equiv \sum_{k=1}^{s} R^{(k)} \qquad (11.13)$$

Variabilitatea intergrupe:

$$V_I = \sum_{k=1}^{s} \frac{\left(T^{(k)}\right)^2}{N_k} - \frac{\left(\sum_{k=1}^{s} T^{(k)}\right)^2}{\sum_{k=1}^{s} N_k} \equiv R - \frac{T^2}{N} \qquad (11.14)$$

Se poate calcula şi variabilitatea totală:

Scanned with OKEN Scanner

$$V_{tot} = \sum_{k=1}^{s} SP^{(k)} - \frac{\left(\sum_{k=1}^{s} T^{(k)}\right)^2}{\sum_{k=1}^{s} N_k} \equiv SP - \frac{T^2}{N} \qquad (11.15)$$

Revenind la exemplul studiat, este utilă organizarea calculelor într-un tabel sintetic de forma următoare:

	Grupul experimental 1	Grupul experimental 2	Grupul experimental 3	Grupul de control	Total
Valori înregistrate	69, 72, 81, 75,70, 76	76, 81, 84, 80,82, 79	80, 88, 91, 79,85	68, 70, 67, 73,71	-
Volumul eşantionului	$N_1=6$	$N_2=6$	$N_3=5$	$N_4=5$	$N=22$
Totalul valorilor	$T^{(1)}=443$	$T^{(2)}=482$	$T^{(3)}=423$	$T^{(4)}=349$	$T=1697$
Raportul $R^{(k)}=\left(T^{(k)}\right)^2/N_k$	$R^{(1)}=32.708,2$	$R^{(2)}=38.720,7$	$R^{(3)}=35.785,8$	$R^{(4)}=24.360,2$	$R=131.574,9$
Suma pătratelor	$SP^{(1)}=32.807$	$SP^{(2)}=38.759$	$SP^{(3)}=35.891$	$SP^{(4)}=24.383$	$SP=131.840$

După completarea tabelului, este facilă determinarea variabilității eroare:

$$V_E = 131.840 - 131.574,9 = 265,1$$

şi a variabilității intergrupe:

$$V_I = 131.574,9 - \frac{1697^2}{22} = 674,5$$

Se determină şi variabilitatea totală şi se verifică justețea calculelor:

$$V_{tot} = 939,6 = 674,5 + 265,1$$

Scanned with OKEN Scanner

Se calculează valorile estimatorilor

$$MS_I = \frac{674,5}{4-1} = 224,8 \quad MS_E = \frac{265,1}{22-4} = 14,7$$

şi valoarea calculată a raportului F:

$$F_{calc} = \frac{224,8}{14,7} = 15,29$$

Valoarea critică se determină din tabelul distribuției F, corespunzător gradelor de libertate

$$l_1 = 4-1 = 3 \text{ şi } l_2 = 22-4 = 18.$$

De exemplu, pentru pragul de semnificație α=0,01 se obține F_{cr}=5,09. Valoarea calculată este mai mare decât cea critică, deci putem accepta ipoteza specifică — mediile grupelor diferă semnificativ ca urmare a influenței exercitate de mărimea investiției.

Intensitatea efectului variabilei independente

Factorul de diferențiere care intervine în problemele de analiză dispersională unifactorială poate fi privit ca o variabilă independentă (*VI*) cu valori diferite pentru fiecare eşantion. Astfel, rezultatul măsurat în urma unui test pe eşantioanele colectivității experimentale se constituie ca variabilă dependentă (*VD*) determinată pe de o parte de valoarea *VI*, dar şi de alți factori aleatori (corespunzători variabilității eroare).

Testul F oferă un indiciu *global* asupra efectului variabilei independente, arătând dacă acest efect este sau nu semnificativ, dar nu permite evaluarea *intensității* acestui efect asupra variabilei dependente. Se poate exprima procentul din variabilitatea (dispersia) valorilor *VD* datorat *VI* utilizând raportul dintre variabilitatea intergrupe şi variabilitatea totală:

$$R_{I.Tot} = \frac{V_I}{V_{tot}} \cdot 100 \qquad (11.6)$$

Valoarea coeficientului $R_{I.Tot}$ este specifică eşantionului studiat şi nu constituie o estimare extensibilă la populația originară. Pentru

Scanned with OKEN Scanner

exemplul studiat $R_{I.Tot}$=71,8%, deci aproximativ 70% din variația venitului net înregistrat în colectivitatea studiată se datorează influenței mărimii investiției, diferită pentru fiecare eşantion.

Teste de comparație între grupe

Asupra mulțimii de *s* grupe independente, după efectuarea analizei dispersionale globale (testul ***F***), se poate recurge la comparații între perechi de grupe sau între submulțimi de mai multe grupe. Aplicarea acestor teste implica grupe de efective egale

$$(N_1 = N_2 = ... = N_s \equiv n).$$

De pildă, în experimentul anterior se studiază efectul administrării de investiții la 3 grupe de studiu, colectivitatea fiind completată cu a 4-a grupă, de control. Rezultatele înregistrate practic au arătat că influența investiției asupra venitului net este semnificativă. În continuare se poate pune problema de a testa efectele investiției comparativ între grupele experimentale luate două câte două sau între ansamblul celor 3 grupe experimentale şi situația — martor (grupa de control). Comparațiile se pot face deci între *perechi de medii* (m_1 cu m_2, m_1 cu m_3 ş.a.m.d.)sau între *combinații de medii* (de exemplu: combinația m_1, m_2, m_3 cu m_4).

Comparația între perechi de medii va fi notată cu variabilă $d_{ij} = m_i - m_j$. Generalizând pentru combinații de medii, comparația poate fi exprimată matematic cu ajutorul unei combinații liniare de medii *sau suma ponderată de medii*, respectând condiția ca suma ponderilor să fie nulă:

$$d = a_1m_1 + a_2m_2 + ... + a_sm_s \equiv \sum_{k=1}^{s}(a_km_k), \qquad (11.17)$$

unde $a_1 + a_2 + ... + a_s = 0$. Convențional, se poate stabili ca grupele care se compară să primească coeficienți sau ponderi pozitive, grupele cu care se compară să aibă ponderi negative, iar mediile care nu apar în diferența respectiva să aibă coeficient zero. De exemplu, la comparația m_1 cu m_2

Scanned with OKEN Scanner

variabila d se va scrie $d_{12} = m_1 - m_2$, observând că grupa care se compară (1) are coeficientul „1" (deci pozitiv), cea cu care se compară (2) are coeficientul „–1" (negativ), iar celelalte medii $(m_3, m_4, \ldots)$ au ponderea „0" şi nu apar în expresie. Suma coeficienţilor este totdeauna zero.

După stabilirea combinaţiei liniare de medii d care urmează a fi supusă testului se calculează o estimare a variantei dintre grupele avute în vedere:

$$MS_{inter}(d) = \frac{d^2}{\frac{1}{n}\sum_{k=1}^{s} a_k^2} \qquad (11.18)$$

Să notam ca acest estimator are totdeauna numărul de grade de libertate egal cu 1 pentru fiecare comparaţie particulară.

Testele de comparaţie între grupe se bazează tot pe distribuţia Fischer — Snedecor. Este necesar deci să calculam valoarea raportului F:

$$F_{calc} = \frac{MS_{inter}(d)}{MS_E}, \qquad (11.19)$$

care urmează a fi comparată cu valoarea critică corespunzătoare gradelor de libertate $l_1 = 1$ si $l_2 \doteq N - s$ şi pragului de semnificaţie α impus.

Regulile de decizie sunt cele specifice testelor bilaterale. Pentru $F_{calc} > Fcr$ diferenţa d între mediile considerate este semnificativă; valorile $F_{calc} < F_{cr}$ indică o diferenţă d nesemnificativă.

Ca tehnică de lucru, se pot face *comparaţii prevăzute* înainte de experiment, caz care implică constituirea grupelor în conformitate cu aceste comparaţii care urmează a fi efectuate după măsurarea rezultatelor. În această situaţie nu este obligatoriu testul global de variantă F, ci se poate trece direct la testele de comparaţie prevăzute.

După realizarea experimentului, unele rezultate particulare (diferenţe înregistrate între anumite grupe) constate pot sugera efectuarea unor *comparaţii aposteriori*. În acest caz, este necesară efectuarea preliminară a testului global şi acceptarea ipotezei inegalităţii mediilor (specifică), după care se poate trece la testele de comparaţie între grupele

Scanned with OKEN Scanner

sau combinațiile de grupe dorite. Pentru testele de comparație se poate impune un prag de semnificație mai sever decât cel folosit la testul global.

Observație: Aceleaşi rezultate se obțin dacă se apelează la comparații între mărimile $T^{(k)}$ (sumele valorilor din grupe) în locul mediilor m_k. Diferențele obținute prin aceste combinații liniare se vor nota cu *D*:

$$D = a_1 T^{(1)} + a_2 T^{(2)} + ... + a_s T^{(s)} \equiv \sum_{k=1}^{s} \left(a_k T^{(k)} \right), \qquad (11.20)$$

unde suma ponderilor $a_1, a_2, ..., a_s$ este totdeauna nula.

În acest caz estimatorul variantei dintre grupe are forma

$$MS_{inter}(D) = \frac{D^2}{n \cdot \sum_{k=1}^{s} a_k^2}, \qquad (11.21)$$

iar raportul ***F*** se calculează similar:

$$F_{calc} = \frac{MS_{inter}(D)}{MS_E}, \qquad (11.22)$$

Gradele de libertate si regulile de decizie rămân neschimbate.

11.3 **Analiza dispersională bifactorială**

În problemele de analiză dispersională unifactorială există un singur factor (variabila independentă) modificat în experiență, efectul acestei modificări determinând legea de variație a măsurătorilor efectuate asupra variabilei dependente. Dacă se introduce un al doilea factor de diferențiere, comportamentul variabilei dependente se va modifica în consecință datorită efectului direct al noii variabile şi a interacțiunii dintre cele doua variabile independente. Problema devine a analiza dispersionala bifactorială.

În problemele de ANOVA bifactorială prezența celor doi factori independenți (numiți generic *A* şi *B*) conduce la posibilitatea formulării şi verificării a 3 ipoteze specifice (cu ipotezele nule aferente):

Scanned with OKEN Scanner

— (H_{1A}): la nivelul populației originare variabila *A* influențează semnificativ variabila dependentă (rezultatul măsurării);

— (H_{1B}): la nivelul populației originare variabila *B* influențează semnificativ variabila dependentă (rezultatul măsurării);

— $(H_{1A\times B})$: la nivelul populației originare există o interacțiune semnificativă a celor doi factori de diferențiere.

Corespunzător, pentru verificarea fiecărei ipoteze se va calcula câte o valoare pentru raportul ***F*** care se compară cu valoarea critică asociată.

Exemplu:

Ne propunem studiul efectului unei investiții asupra venitului net. În acelaşi timp, ne interesează influența pregătirii manageriale asupra rezultatelor, precum şi eventuala interacțiune dintre cei doi factori care condiționează venitul net.

Să notăm cu *A* volumul investiției, cu una din modalitățile „mică", „medie" şi „avansată" — a_1, a_2 respectiv a_3. Pentru variabila independentă *B*, natura materialului de reținut, vom reține două modalități: „bună" şi „foarte bună", notate b_1, respectiv b_2. Combinația dintre cele două variabile conduce la structurarea colectivității testate în $3\times 2 = 6$ grupuri.

Rezultatele obținute după o anumită perioadă sunt prezentate în tabelul următor:

	b_1	b_2	Total
a_1	7, 9, 8, 10, 10, 7, 8 $N_{a_1b_1}$ =7 $T_{a_1b_1}$ =59 , $SP_{a_1b_1}$ =507 $T^2_{a_1b_1}/N_{a_1b_1}$ =497,28	9, 8, 8, 10, 7, 7, 8, 6 $N_{a_1b_2}$ =8 $T_{a_1b_2}$ =63 , $SP_{a_1b_2}$ =507 $T^2_{a_1b_2}/N_{a_1b_2}$ =496 ,13	N_{a_1} =15 T_{a_1} =122 R_{a_1} =992,27

Scanned with OKEN Scanner

a_2	6, 8, 8, 7, 8, 7, 6 $N_{a_2b_1}$ =7 $T_{a_2b_1}$ =50 , $SP_{a_2b_1}$ =362 $T^2_{a_2b_1} / N_{a_2b_1}$ =357 ,14	8, 7, 7, 8, 6, 6, 7, 5 $N_{a_2b_2}$ =8 $T_{a_2b_2}$ =54 , $SP_{a_2b_2}$ =372 $T^2_{a_2b_2} / N_{a_2b_2}$ =36 4,50	N_{a_2} =15 T_{a_2} =104 R_{a_2} =721,07
a_3	5, 6, 7, 5, 6, 6, 4 $N_{a_3b_1}$ =7 $T_{a_3b_1}$ =39 , $SP_{a_3b_1}$ =223 $T^2_{a_3b_1} / N_{a_3b_1}$ =217, 28	6, 6, 5, 5, 6, 4, 5, 4 $N_{a_3b_2}$ =8 $T_{a_3b_2}$ =41 , $SP_{a_3b_2}$ =215 $T^2_{a_3b_2} / N_{a_3b_2}$ =21 0,13	N_{a_3} =15 T_{a_3} =80 R_{a_3} =426,67
Total	N_{b_1} =21 T_{b_1} =148 R_{b_1} =1043,05	N_{b_2} =24 T_{b_2} =158 R_{b_2} =1040,17	N =45 T =306 R =2142,46 SP =2186

Vom nota cu s_A, s_B numărul de categorii al variabilei independente *A*, respectiv *B*. În problemele ANOVA bifactoriale trebuiesc calculate următoarele valori:

— suma valorilor din fiecare eşantion (identificat prin *k* si *l*, indecşi corespunzători modalităţilor celor doua variabile independente):

$$T_{kl} = \sum_{i=1}^{N_{kl}} X_i^{(kl)} \quad , \; k = 1,2,...s_A \; , l = 1,2,...s_B \qquad (11.23)$$

— suma pătratelor valorilor din fiecare eşantion:

Scanned with OKEN Scanner

$$SP_{kl} = \sum_{i=1}^{N_{kl}} \left(X_i^{(kl)}\right)^2 \quad , \; k = 1,2,...s_A \; , l = 1,2,...s_B \tag{11.24}$$

— numărul de subiecţi incluşi în fiecare categorie a celor două variabile:

$$N_k = \sum_{l=1}^{s_B} N_{kl} \quad , k = 1,2,...s_A \quad \text{, respectiv} \tag{11.25}$$

$$N_l = \sum_{k=1}^{s_A} N_{kl} \quad , l = 1,2,...s_B$$

— totalul valorilor corespunzătoare categoriilor variabilelor:

$$T_k = \sum_{l=1}^{s_B} T_{kl} \quad , k = 1,2,...s_A \quad \text{, respectiv} \tag{11.26}$$

$$T_l = \sum_{k=1}^{s_A} T_{kl} \quad , l = 1,2,...s_B$$

— raportul $R = T^2/N$ pentru fiecare categorie a celor două variabile:

$$R_k = {}^{T_k^2}\!/_{N_k} \quad , k = 1,2,...s_A \quad \text{, respectiv} \tag{11.27}$$

$$R_l = {}^{T_l^2}\!/_{N_l} \quad , l = 1,2,...s_B$$

— numărul total de subiecţi, totalul general, raportul *R* şi suma pătratelor *SP* pentru întreaga colectivitate:

$$N = \sum_{k=1}^{s_A}\left(\sum_{l=1}^{s_B} N_{kl}\right) \equiv \sum_{k=1}^{s_A} N_k \equiv \sum_{l=1}^{s_B} N_l$$

$$T = \sum_{k=1}^{s_A}\left(\sum_{l=1}^{s_B} T_{kl}\right) \equiv \sum_{k=1}^{s_A} T_k \equiv \sum_{l=1}^{s_B} T_l \tag{11.28}$$

$$R = \sum_{k=1}^{s_A}\left[\sum_{l=1}^{s_B}\left({}^{T_{kl}^2}\!/_{N_{kl}}\right)\right]$$

Scanned with OKEN Scanner

$$SP = \sum_{k=1}^{s_A} \left(\sum_{l=1}^{s_B} SP_{kl} \right)$$

Variabilitatea eroare (intragrupe) se calculează cu formula:

$$V_E = SP - R = 2186 - 2142,46 = 43,54$$

Variabilitatea intergrupe indusă de variabila independentă *A*:

$$V_A = \sum_{k=1}^{s_A} R_k - \frac{T^2}{N} = 992,27 + 721,07 + 426,67 - \frac{306^2}{45} = 59,21$$

Variabilitatea intergrupe indusă de variabila independentă *B*:

$$V_B = \sum_{l=1}^{s_B} R_l - \frac{T^2}{N} = 1043,05 + 1040,17 - \frac{306^2}{45} = 2,41$$

Variabilitatea intergrupe determinată de interacțiunea dintre variabilele *A* şi *B*:

$$V_{A\times B} = R - \frac{T^2}{N} - V_A - V_B = 2142,46 - \frac{306^2}{45} - 59,21 - 2,41 = 0,04$$

Variabilitatea totală:

$$V_{tot} = SP - \frac{T^2}{N} = 2186 - \frac{306^2}{45} = 105,2$$

Variabilitatea totală reprezintă suma corespunzătoare celor 4 surse de dispersie a datelor. Se verifică astfel justețea calculelor:

$$V_{tot} = V_E + V_A + V_B + V_{A\times B} = 43,54 + 59,21 + 2,41 + 0,04 = 105,2$$

Se determină estimatorii corespunzători celor 4 variante. Gradele de libertate corespund valorilor de la numitorul fracțiilor:

$$MS_E = \frac{V_E}{N - s_A \cdot s_B} = \frac{43,54}{45 - 3 \cdot 2} = 1,12$$

$$MS_A = \frac{V_A}{s_A - 1} = \frac{59,21}{3 - 1} = 29,61$$

Scanned with OKEN Scanner

$$MS_B = \frac{V_B}{s_B - 1} = \frac{2,41}{2-1} = 2,41$$

$$MS_{A\times B} = \frac{V_{A\times B}}{(s_A - 1)(s_B - 1)} = \frac{0,04}{(3-1)(2-1)} = 0,02$$

şi se determină valorile rapoartelor F:

$$F_{calc_A} = \frac{MS_A}{MS_E} = \frac{29,61}{1,12} = 26,43$$

$$F_{calc_B} = \frac{MS_B}{MS_E} = \frac{2,41}{1,12} = 2,15$$

$$F_{calc_{A\times B}} = \frac{MS_{A\times B}}{MS_E} = \frac{0,02}{1,12} = 0,018$$

Valorile critice corespunzătoare pragului de semnificație α =0,05:

$F_{cr_A}\Big|_{\substack{l_1=2\\ l_2=39\\ \alpha=0,05}} \cong 3,23$, $F_{cr_B}\Big|_{\substack{l_1=1\\ l_2=39\\ \alpha=0,05}} \cong 4,08$, $F_{cr_{A\times B}}\Big|_{\substack{l_1=2\\ l_2=39\\ \alpha=0,05}} \cong 3,23$

Regulile de decizie sunt aceleaşi ca şi în cazul problemei ANOVA unifactoriale (specifice testelor bilaterale). Observăm ca doar $F_{calc\,A}$ are o valoare semnificativă (mai mare decât valoarea critică corespunzătoare), deci la nivelul populației valoarea investiției are o influență semnificativă asupra venitului net. Influența pregătirii manageriale, precum şi interacțiunea dintre cei doi factori nu sunt validate, deoarece valorile calculate sunt mai mici decât valorile critice corespunzătoare (deci nu se acceptă ipotezele specifice).

11.4. **Aplicații**

A1. Se compară trei metode diferite de laborator ce urmăresc determi-narea concentrării de bacterii în apa potabilă. Probele conduc la următoarele rezultate:

Scanned with OKEN Scanner

Metoda	I	II	II
Selecţii	1,86	1,90	1,96
	1,92	1,84	2,03
	1,78		1,98

K=3

- Verificăm ipoteza H_o: nu există diferenţă între eficienţele celor trei metode, deci:

$$H_o : m_1 = m_2 = m_3$$

Calculăm mediile de selecţie ale celor 3 probe:

$$\bar{x}_{1*} = \frac{1,86 + 1,92 + 1,78}{3} = 1,85 \qquad n_1 = 3$$

$$\bar{x}_{2*} = \frac{1,90 + 1,84}{2} = 1,87 \qquad n_2 = 2$$

$$\bar{x}_{3*} = \frac{1,96 + 2,03 + 1,98}{3} = 1,99 \qquad n_3 = 3$$

$$n = \sum_{i=1}^{k} n_i = 8$$

$$\bar{\bar{x}} = \frac{1}{n}\sum_{i=1}^{3} n_i \bar{x}_{i*} = \frac{1}{8}(3\cdot 1,85 + 2\cdot 1,87 + 3 1,99) \Rightarrow$$

$$\bar{\bar{x}} = \frac{1}{8}\cdot(5,55 + 3,74 + 5,97) = \frac{1}{8}\cdot 15,26 = 1,91$$

Calculăm valorile sumelor S_1, S_{12}:

S_1 = variaţia factorială $\left(S_1 = \sum_{i=1}^{k} n_i \left(\bar{x}_{i*} - \bar{\bar{x}}\right)^2 \right.$

$$S_1 = \left[n_1\left(\bar{x}_{1*} - \bar{\bar{x}}\right)^2 + n_2\left(\bar{x}_{2*} - \bar{\bar{x}}\right)^2 + n_3\left(\bar{x}_{3*} - \bar{\bar{x}}\right)^2 \right] =$$

$$= 3\cdot(1,85 - 1,91)^2 + 2(1,87 - 1,91)^2 + 3\,(1,99 - 1,91)^2 =$$

$$= \underbrace{3\cdot 0,0036}_{0,0108} + \underbrace{2\cdot 0,0016}_{0,0032} + \underbrace{3\cdot 0,0064}_{0,0192} = 0,0332$$

Scanned with OKEN Scanner

$$S_{12} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{i*}\right)^2 (1{,}86-1{,}85)^2 + (1{,}92-1{,}85)^2 + (1{,}78-1{,}85)^2 +$$

$$+ (1{,}90-1{,}87)^2 + (1{,}84-1{,}87)^2 + (1{,}96-1{,}99)^2 + (2{,}03-1{,}99)^2 +$$

$$+ (1{,}98-1{,}99)^2 = 0{,}0001 + 0{,}0049 + 0{,}0049 + 0{,}0009 + 0{,}0009 + 0{,}0009 +$$

$$0{,}0016 + +0{,}0001 = 0{,}0143$$

S_{12} caracterizează abaterile interne ale selecțiilor k=3; n=8

$$Z = \frac{S_1}{2} : \frac{S_{12}}{5} = \frac{0{,}0336}{2} : \frac{0{,}0143}{5} = \frac{0{,}0336}{2} \cdot \frac{5}{0{,}0143} \Rightarrow Z = 5{,}80$$

Utilizăm tabelele pentru repartiția Fischer

$$\left.\begin{array}{l} F\,(k\text{-}1,\ n\text{-}k) = F\,(2,5) \text{ cu } \alpha = 0{,}01 \end{array}\right\} \Rightarrow F_{0{,}99}\ (2,5)$$

k-1, n-k = sunt grade de libertate

$F_{0,99}$ (2,5) = 13,27.

- Deoarece: $Z < F_{0,99}$ (2,5) (5,8 < 13,27) se acceptă ipoteza H_0, adică cele 3 metode sunt la fel de eficiente.

A.2. Să se prezinte analiza dispersională cu doi factori fără interacție între ei

Răspuns:

Populația supusă cercetării *suferă două influențe* datorită acțiunii asupra ei a doi factori:

- factorul A – cu K nivele şi
- factorul B – cu l nivele

Ipoteze:

1. x_{ij}, $1 \le i \le k$; $1 \le j \le l$ sunt variabile aleatoare independente repartizate $N(m_{ij}, \sigma^2)$

Scanned with OKEN Scanner

2. Există o singură *valoare de selecție* pentru fiecare pereche (i, j), notată cu x_{ij}.
3. Nu există interacție între cei doi factori, deci efectele acestora se însumează, adică $m_{ij} = m + (m_{I*} - m) + (m_{*j} - m)$ unde: parametrul „m" reprezintă aici valoarea medie calculată (luând în considerație influențele tuturor nivelelor celor doi factori)

$$m_{i*} = \frac{1}{l}\sum_{j=1}^{l} m_{ij}$$

$$m_{*j} = \frac{1}{k}\sum_{i=1}^{k} m_{ij}$$

$$m = \frac{1}{kl}\sum_{i=1}^{k}\sum_{j=1}^{l} m_{ij}$$

Notăm :

$\alpha_I = m_{I*} - m$ $\qquad i = \overline{1,k}$

$\beta_j = m_{*j} - m$ $\qquad j = \overline{1,l}$

α_i – *reprezintă efectul factorului A la nivelul i*

β_j – *reprezintă efectul factorului B la nivelul j.*

A.3. **Să se verifice ipoteza H_o, cu datele din A.2.**

$H_o : m_{ij} = m, \quad i = \overline{1,k}, \quad j = \overline{1,l}$, față de alternativa H_1,

H_1 : cel puțin o valoare $m_{ij} \neq m$

Răspuns :

1. H_o este adevărată dacă $\alpha_I = 0$

$(\forall) i = \overline{1,k}$ și $\beta_j = 0$ pentru orice $j = \overline{1,l}$

- **Verificarea ipotezei H_0** se transformă în verificarea a două ipoteze :

$$H_{01} : \sum_{i=1}^{k} \alpha_i^2 = 0 \text{ și}$$

Scanned with OKEN Scanner

$$H_{02}: \sum_{j=1}^{l} \beta_j^2 = 0$$

$$\left| m_{ij} - m = \left(m_{ij} - m_{ix} - m_{xj} - m\right) + \left(m_{ix} - m\right) + \left(m_{xj} - m\right) \right|$$

Estimațiile parametrilor

Notații.

$$S = \sum_{j=1}^{l} \sum_{i=1}^{k} \left(x_{ij} - \bar{\bar{x}}\right)^2$$

$$S_1 = \sum_{j=1}^{l} \sum_{i=1}^{k} \left(\bar{x}_{ij} - \bar{\bar{x}}\right)^2$$

$$S_2 = \sum_{j=1}^{l} \sum_{i=1}^{k} \left(\bar{x}_{*j} - \bar{\bar{x}}\right)^2$$

$$S_{12} = \sum_{j=1}^{l} \sum_{i=1}^{k} \left(x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{\bar{x}}\right)^2$$

A.4.

1. Verificarea ipotezei $H_{01}: \sum_{i=1}^{k} \alpha_1^2 = 0$ față de alternativa H_{01} : „cel puțin o valoare $\alpha_i \neq 0$ “, cu date din A.3.

Răspunsuri:

- Această ipoteză revine de fapt la ipoteza:

$$m_{1*} = m_{2*} = \cdots = m_{k*} = m$$

ceea ce conduce la utilizarea statisticii test $\frac{S_1}{K-1} : \frac{S_{12}}{(k-1)(l-1)}$

Regiunea critică este (c, ∞) unde C este definită de

$$0{,}01 = P\left(\frac{S_1}{k-1} : \frac{S_{12}}{(k-1)(l-1)} > C \right)$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Estimaţiile parametrilor

$$\bar{x}_{i*} = \frac{1}{l}\sum_{j=1}^{l} X_{ij}; \quad \bar{x}_{*j} = \frac{1}{k}\sum_{i=1}^{k} X_{ij}; \quad \bar{\bar{x}}_{..} = \frac{1}{l \cdot k}\sum_{i=1}^{k}\sum_{j=1}^{l} X_{ij}$$

- Estimaţiile de maximă verosimilitate pentru parametri: m, α_i şi β_j sunt:

$$\hat{m} = \bar{\bar{X}}$$
$$\hat{\alpha}_i = \bar{x}_{i*} - \bar{\bar{x}}$$
$$\hat{\beta}_j = \bar{x}_{*j} - \bar{\bar{x}}$$

- *Pentru valorile de selecţie avem identitatea:*

$X_{ij} - \bar{\bar{x}} = \left(x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{\bar{x}}\right) + \left(\bar{x}_{i*} - \bar{\bar{x}}\right) + \left(\bar{x}_{*j} - \bar{\bar{x}}\right)$ care, *ridicată la pătrat* şi în condiţiile în care nu există interacţie între cei doi factori devine:

$$\sum_{j=1}^{l}\sum_{i=1}^{k}\left(X_{ij} - \bar{\bar{X}}\right)^2 = \sum_{j=1}^{l}\sum_{i=1}^{k}\left(\bar{x}_{j*} - \bar{\bar{x}}\right)^2 + \sum_{j=1}^{l}\sum_{i=1}^{k}\left(x_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{\bar{x}}\right)^2$$

2. Verificarea ipotezei H_{02},

$H_{02} : \sum_{j=1}^{l} \beta_j^2 = 0$ faţă de alternativa

H_{21} : „cel puţin o valoare $\beta_j \neq 0$“

Observaţie : În cazul ipotezei H_{02},

H_{02}: $m_{*1} = m_{*2} = \ldots = m_{*l} = m$

statistica test folosită va fi :

$$F = \frac{S_2}{l-1} : \frac{S_{12}}{(k-1)(l-1)}$$

În acest caz, regiunea critică (C, ∞) unde C este definită de :

$$0{,}01 = P\,(F > C)$$

SURSA VARIAȚIEI	Sumele de pătrate	Grade de libertate	Statistica test folosită
Între nivelele factorului B	S_1	k – 1	$f_1 = \frac{S_1}{k-1} : \frac{S_{12}}{(k-1)(l-1)}$
Între nivelele factorului A	S_2	l – 1	$f_2 = \frac{S_2}{l-1} : \frac{S_{12}}{(k-1)(l-1)}$
Eroare Total	S_{12} s	(k–1) (l–1) k l – 1	

Astfel ipoteza : $H_0 : m_{ij} = m \ (\forall) \ i = \overline{1,k}, \ j = \overline{1,l}$ se descompune în verificarea a două ipoteze :

- $H_{01} : \alpha_1 = \alpha_2 = \ldots \alpha_k = 0$ şi
- $H_{02} : \beta_1 = \beta_2 = \ldots = \beta_l = 0$

A.5.

Se consideră datele din următorul tabel.

		FACTORUL B					
	i / j	1	2	3	4	5	
Factorul A	1	45,3	42,2	39,6	26,8	45,8	$\overline{X}_{*1}$
	2	44,1	44,1	38,4	38	47,2	$\overline{X}_{*2}$
	3	42,7	42,7	42,6	42,2	46,8	$\overline{X}_{*3}$
	4	43,5	45,8	47,9	37,8	46,2	$\overline{X}_{*4}$
		$\overline{X}_{*1}$	$\overline{X}_{*2}$	$\overline{X}_{*3}$	$\overline{X}_{*4}$	$\overline{X}_{*5}$	

Factorul B reprezintă 5 subiecți (l=5) aflați sub acțiunea factorului A (cu patru trepte) (k = 4). Considerând că valorile din tabel reprezintă valori caracteristice să se efectueze analiza dispersională (cu doi factori fără interacție între ei)

Scanned with OKEN Scanner

$$\overline{X}_{1*} = \frac{1}{l}\sum_{j=1}^{l}\overline{X}_{1j};\quad \overline{X}_{2*} = \frac{1}{l}\sum_{j=1}^{l}X_{2j};\quad \overline{X}_{3*} = \frac{1}{l}\sum_{j=1}^{l}X_{3j};\quad \overline{X}_{4*} = \frac{1}{l}\sum_{j=1}^{l}X_{4j}$$

Răspuns:

$\overline{x}_{1*} = 41,7;\quad \overline{x}_{2*} = 42,36;\quad \overline{x}_{3*} = 43,4;\quad \overline{x}_{4*} = 44,24$

$$\overline{x}_{*1} = \frac{1}{k}\sum_{i=1}^{k}X_{i1} = 43,9;\quad \overline{x}_{*2} = 43,7;\quad \overline{x}_{*3} = 42,12$$

$\overline{x}_{*4} = 36,2;\quad \overline{x}_{*5} = 46,5$

$$\overline{\overline{X}} = \frac{1}{lk}\sum_{i=1}^{k}\sum_{j=1}^{l}X_{ij} = 42,48$$

$$S_1 = \sum_{j=1}^{l}\sum_{i=1}^{k}\left(\overline{x}_{i*} - \overline{\overline{x}}\right)^2 = 5\sum_{i=1}^{k}\left(\overline{x}_{i*} - \overline{\overline{x}}\right)^2 = 22,8$$

$$S_2 = \sum_{j=1}^{l}\sum_{i=1}^{k}\left(\overline{x}_{*j} - \overline{\overline{x}}\right)^2 = 4\sum_{j=1}^{l}\left(\overline{x}_{*j} - \overline{\overline{x}}\right)^2 = 236,96$$

$k = 4,\qquad l = 5$

$$S_{12} = \sum_{j=1}^{l}\sum_{i=1}^{k}\left(x_{ij} - \overline{x}_{i*} - \overline{x}_{*J} + \overline{\overline{x}}\right)^2 = 273,01$$

Calculăm f_1 şi f_2

$$f_1 = \frac{S_1}{k-1} : \frac{S_{12}}{(k-1)(l-1)} = \frac{22,8}{3\cdot 4} : \frac{273,01}{3\cdot 4} = \frac{22,8}{3} \cdot \frac{3\cdot 4}{273,01} = 0,3341$$

$$f_2 = \frac{S_2}{l-1} : \frac{S_{12}}{(k-1)(l-1)} = \frac{236,96}{4} : \frac{273,01}{3\cdot 4} = \frac{236,96}{4} \cdot \frac{3\cdot 4}{273,01} = 2,60$$

Verificăm ipoteza H_{01}: $m_{1*} = m_{2*} = m_{3*} = m_{4*} = m \left(\Leftrightarrow H_{01} : \sum_{i)1}^{k} \alpha_i^2 = 0 \right)$

Utilizăm în acest caz statistica f_1

Scanned with OKEN Scanner

Regiunea critică (C, ∞); Unde C este definită de $f(n, m)$

$0{,}01 = P(f_1 > C)$

$f_{0,99}(k-1, (k-1)(l-1) = f_{0,99}(3; 12) = 5{,}95$

deci regiunea critică este $(595, +\infty)$

Cum valoarea statisticii $f_1 \notin (5{,}95; +\infty)$

se acceptă ipoteza H_{01}

Se utilizează şi statistica f_2 pentru verificarea ipotezei H_{02}:

$$m_{*1} = m_{*2} = \ldots = m_{*5}$$

Regiunea critică $(C, +\infty)$

se determină din condiția:

$$0{,}01 = P(f_2 > C) \text{ sau } 0{,}99 = P(f_2 < C)$$

$f_2 = f_{0,99}(l-1; (k-1)(l-1) = f_{0,99}(4,12) = 5{,}41 \Rightarrow$ regiunea critică $(5{,}41; +\infty)$ şi cum $f_2 \notin (5{,}41; +\infty)$ se acceptă H_{02}.

Concluzie: Din datele experimentale prezentate putem accepta ipoteza de egalitate a mediilor.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 12

ANALIZA CLUSTER[1]

12.1. Noțiuni generale

Analiza „CLUSTER" este o metodă de grupare a elementelor unei mulțimi, în submulțimi, pe baza uneia sau mai multor însuşiri ale elementelor mulțimii. Cuvântul cluster are semnificația de a strânge în grupe, de a clasifica sau de rezultat al grupării. Elementele care sunt grupate prin analiză cluster sunt de regulă, cazuri statistice, dar pot fi şi variabile.

Algoritmul metodelor cluster utilizează noțiunea de *„distanță dintre elemente"*, prin care caracterizăm două elemente ale mulțimii. La fiecare pas al analizei sunt reuniți doi clusteri în unul singur. În final se obține un singur cluster care coincide cu mulțimea elementelor (considerată inițial). Modelele cluster pot să difere în funcție de modul de definire a distanței, de procedeul de grupare şi de criteriul care decide nivelul optim de grupare.

1. **Noțiunea de grupă** (cluster, categorie)

Se consideră zece fire: A,B,C,D,E,F,G,H,I,J, caracterizate fiecare de patru variabile: cifra de afaceri (C.A. – mld. lei); profitul net (P.N. — mld. lei); productivitate (P zeci de mii lei/min.); nivelul cheltuielilor de reclamă şi publicitate (mil. lei), la nivelul anului 2000.

[1] * cluster = grup, mulțime, grămadă, a forma un grup.

Nr. crt.	Firma	C.A.	P.N.	Prod.	R.P. (reclamă şi publicitate)
1	A	30	11	48	5
2	B	26	5	10	3
3	C	45	16	110	3
4	D	24	9	65	5
5	E	43	15	100	3
6	F	26	6	10	5
7	G	30	7	35	5
8	H	15	6	5	8
9	I	18	7	30	4
10	J	22	7	40	5
Medie		27,90	8,90	45,30	4,60
Dispersie		9,2244	3,6729	34,7248	1,4283

Separăm două firme (C) şi (E), care au cifra de afaceri şi productivitate, cele mai mari, într-o grupă separată pentru care întocmim un tabel asemănător.

Nr.crt.	Firma	C.A.	P.N.	Prod.	R.P. (reclamă şi publicitate
1	C	45	16	110	3
2	E	43	15	100	3
Media		44,00	15,50	105,00	3,00
Dispersia		1,000	0,5000	5,0000	0,0000

Se observă că dispersiile în cadrul grupei formată din două firme cu performanţe apropiate sunt mai mici decât pentru întreaga populaţie considerată iniţial.
Să adaugăm la cele două firme C, E şi firmele B, F, G într-un nou tabel:

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Nr.crt.	Firma	C.A.	P.N.	Prod.	R.P.
1	C	45	16	110	3
2	E	43	15	100	3
3	B	26	5	10	3
4	F	26	6	10	5
5	G	30	7	35	5
Media		22,3750	9,80	53,00	3,80
Dispersie		16,3855	4,7074	43,5431	0,9798

Se constată că o dată cu creşterea numărului elementelor dintr-o grupă, creşte şi dispersia valorilor variabilelor studiate.

Dacă luăm în considerare numai prima variabilă (cifra de afaeri), observăm că cea mai mare asemănare există între firmele B şi F care au valori identice ($CA_B = CA_F = 26$) ceea ce înseamnă dispersie zero. Adăugând în continuare câte o firmă, cu valoarea cifrei de afaceri apropiată de a firmelor două, dispersia creşte progresiv. Se continuă alegerea de firme până când le luăm pe toate într-o singură grupă.

Presupunem că ne oprim la un grup de patru firme. Formulăm următoarea problemă: care anume grupă de patru firme are dispersia minimă? Se mai pune şi problema identificării a două firme cu care începe analiza cluster. Soluţia constă în elaborarea metodei de analiză cluster.

2. Noţiunea de distanţă

Obţinerea de clusteri compacţi (cu elemente asemănătoare) se bazează pe o funcţie „distanţă" definită pe mulţimea perechilor de elemente.

Funcţia „distanţă" d(x,y) îndeplineşte următoarele condiţii :

1. $d(x,y) > 0$
2. $d(x,y) = 0$ dacă şi numai dacă $x = y$ (12.1)
3. $d(x,y) \leq d(x,z) + d(z,y)$.

Axiomele 1,2,3 sunt cunoscute şi sub denumirea de axiomele distanţei.

În analiza cluster se folosesc următoarele tipuri de distanţe

- SEUCLID — pătratul distanței euclidiene

$$d(X,Y) = \sum_{i=1}^{m}(X_i - Y_i)^2 \qquad (12.2)$$

- EUCLID — distanța euclidiană

$$d(X,Y) = \sqrt{\sum_{=1}^{n}(X_i - Y_i)^2} \qquad (12.3)$$

- BLOC ((Manhattan) — suma abaterilor

$$d(X,Y) = \sum_{i=1}^{m}|X_i - Y_i| \qquad (12.4)$$

- THE BYCHEV — abaterea maximă

$$d(X,Y) = \max|X_i - Y_i| \qquad (12.5)$$

- MINKOWSKI — distanță euclidiană generalizată

$$d(X,Y) = \left(\sum_{i=1}^{m}|X_i - Y_i|^p\right)^{\frac{1}{p}} \qquad (12.6)$$

pentru p = 2 se obține distanța euclidiană

- POWER (p, r) — distanța Minkowski generalizată

$$d(X,Y) = \left(\sum_{i=1}^{m}|X_i - Y_i|^p\right)^{\frac{1}{r}} \qquad (12.7)$$

În toate aceste definiții contează unitatea de măsură a variabilelor care intervin. Pentru a diminua influența unităților de măsură, variabilele sunt transformate în variabile standard (Z) cu media zero şi dispersia unu.

Observație: Există posibilitatea de a defini mărimi care nu verifică axiomele distanței, acestea fiind definite astfel încât cea mai mică diferențiere (cea mai mare asemănare) se obține pentru o valoare maximă a mărimii calculate.

Cele mai reprezentative mărimii de acest tip, numite „similitudini" sunt:

Scanned with OKEN Scanner

- COSINE — cosinusul vectorilor X şi Y

$$C(X,Y)=\frac{\sum_{i=1}^{m} X_i Y_i}{\sqrt{\left(\sum_{i=1}^{m} X_i^2\right)\cdot\left(\sum_{i=1}^{m} Y_i^2\right)}} \qquad (12.8)$$

- CORELAȚIA PEARSON — corelația vectorilor X şi Y

$$S(X,Y)=\frac{\sum_{i=1}^{m} Zx_i \cdot Zy_i}{n-1} \qquad (12.9)$$

unde Z_{xi} sunt valorilor standardizate Z score a variabilei X; n fiind numărul vectorilor. Aceste două mărimi se folosesc în cazul elementelor variabile.

3. **Distanța dintre clusteri**

Algoritmul utilizat în analiza cluster foloseşte noțiunea de distanță dintre clusteri, care se defineşte cu ajutorul distanței dintre elemente.

Se utilizează următoarele distanțe între mulțimi de elemente:

- SINGLE (Single linkage or nearest neighbor) — cea mai mică dintre distanță dintre elementele celor două grupe:

$$D(\Phi,\Psi) = \min d(X,Y); \quad X\in\emptyset,\ Y\in\Psi \qquad (12.10)$$

- COMPLETE (Complete linkage or furthest neighbor) — cea mai mare dintre distanțele dintre elementele celor două grupe:

$$D(\Phi,\Psi) = \max d(X,Y) \qquad (12.11)$$

- BAVERAGE (Average linkage betmeen groups, UPGMA) — media distanțelor dintre elementele celor două grupe:

$$D(F,Z)=\frac{1}{N_\phi \cdot N_\psi}\cdot\sum_{x\in\phi}\sum_{y\in\psi} d(X,Y) \qquad (12.12)$$

- WAVERRAGE (Average linkage within groups) — se calculează media distanțelor dintre elementele posibilului cluster reunit pentru

Scanned with OKEN Scanner

fiecare pereche de clusteri existenţi la acel moment şi se alege perechea pentru care această medie este cea mai mică:

$$D(\Phi,\Psi)=\frac{1}{(N_\phi+N_\psi)(N_\phi+N_\psi-1)}\sum_{\substack{x,y\in\phi\vee\psi\\ x\neq y}} d(x,y); \qquad (12.13)$$

- WARD (Word's met hod) — utilizează numai distanţa euclidiană la pătrat; se calculează pentru fiecare cluster media fiecărei variabile, după care distanţa dintre două clustere este media distanţelor de la elementul mediu la toate elementele celuilalt cluster;
- CENTROID (Controid clustering, UPGMC) — distanţa dintre mediile corespunzătoare celor doi clusteri;
- MEDIAN (Median clustering, WPGMC) — distanţa dintre mediile corespunzătoare celor doi clusteri, dar la alegerea clusterilor ce se unesc se ţine seama de mărimea lor (de numărul de elementele din clusteri).

A.4.

Se aplică algoritmul pentru exemplul cu cele zece firme în care variabilele (cifra de afaceri — C.A. : profitul net — P.N. — productivitatea — P.I.: nivelul cheltuielilor de reclamă şi publicitatea — R.P.) se transformă în variabile standard Z SCORE, notate ZCA, ZPN, ZP, ZRP

$$\left(Z=\frac{x-\overline{x}_a}{\delta}\right)$$

Firma	ZCA	ZPN	ZP	ZRP
A	0,6914	1,0957	0,4582	0,3347
B	–0,6256	–2,0350	–5,9904	–1,3388
C	5,6302	3,7047	10,9795	–1,3388
D	–1.2841	0,0522	3,3431	0,3347
E	4,9717	3,1829	9,2825	–1,3388
F	–0,6256	–1,5132	–5,9904	0,3347
G	0,6914	– 0,9914	– 1,7479	0,3347
H	– 4,2473	– 1,5132	– 6,8389	4,7871
I	– 3,2596	– 0,9914	– 2,5964	– 0,5021
J	– 1,9426	– 0,9914	– 0,8994	0,3347

Scanned with OKEN Scanner

$\overline{X}_{CA}$= 27,90; $\overline{X}_{PN}$=8,90; $\overline{X}_{P}$ = 45,30; $\overline{X}_{RP}$ = 4,60

σ_{CA} = 3,0372; σ_{PN} = 1,9165; σ_{P} = 5,8928; σ_{RP} = 1,1951

Distanța dintre elemente se alege, ca fiind, pătratul distanței euclidiene (Squared Euclidean distance).

Pentru exemplificare calculăm pătratul distanței dintre fişierele C şi D:

Scanned with OKEN Scanner

12.2. Program CLUSTER
ANALIZA CLUSTER

CA=cifra de afaceri , PN=profit net ,P=productivitate , RP=reclama si publicitate

$n := 10 \quad m := 4 \quad i := 1..n \quad j := 1..m \quad k := 1..n$

ORIGIN ≡ 1

$$T = \begin{pmatrix} \text{FIRMA} & \text{CA} & \text{PN} & \text{P} & \text{RP} \\ \text{A} & 30 & 11 & 48 & 5 \\ \text{B} & 26 & 5 & 10 & 3 \\ \text{C} & 45 & 16 & 110 & 3 \\ \text{D} & 24 & 9 & 65 & 3 \\ \text{E} & 43 & 15 & 100 & 3 \\ \text{F} & 26 & 6 & 10 & 5 \\ \text{G} & 30 & 7 & 35 & 5 \\ \text{H} & 15 & 6 & 5 & 8 \\ \text{I} & 18 & 7 & 30 & 4 \\ \text{J} & 22 & 7 & 40 & 5 \end{pmatrix} \qquad T1 := \begin{pmatrix} 30 & 11 & 48 & 5 \\ 26 & 5 & 10 & 3 \\ 45 & 16 & 110 & 3 \\ 24 & 9 & 65 & 3 \\ 43 & 15 & 100 & 3 \\ 26 & 6 & 10 & 5 \\ 30 & 7 & 35 & 5 \\ 15 & 6 & 5 & 8 \\ 18 & 7 & 30 & 4 \\ 22 & 7 & 40 & 5 \end{pmatrix}$$

$T11 := T1^{\langle 1 \rangle} \quad T12 := T1^{\langle 2 \rangle} \quad T13 := T1^{\langle 3 \rangle} \quad T14 := T1^{\langle 4 \rangle}$

$m1 := \text{mean}(T11)$	$\sigma 1 := \text{stdev}(T11)$	$m1 = 27.9$	$\sigma 1 = 9.2244$
$m2 := \text{mean}(T12)$	$\sigma 2 := \text{stdev}(T12)$	$m2 = 8.9$	$\sigma 2 = 3.6729$
$m3 := \text{mean}(T13)$	$\sigma 3 := \text{stdev}(T13)$	$m3 = 45.3$	$\sigma 3 = 34.7248$
$m4 := \text{mean}(T14)$	$\sigma 4 := \text{stdev}(T14)$	$m4 = 4.4$	$\sigma 4 = 1.4967$

$$Z0_{i,1} := \frac{T1_{i,1} - m1}{\sigma 1} \qquad Z0_{i,2} := \frac{T1_{i,2} - m2}{\sigma 2} \qquad Z0_{i,3} := \frac{T1_{i,3} - m3}{\sigma 3} \qquad Z0_{i,4} := \frac{T1_{i,4} - m4}{\sigma 4}$$

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Z0 =

	1	2	3	4
1	0.2277	0.5718	0.0778	0.4009
2	-0.206	-1.0618	-1.0166	-0.9354
3	1.8538	1.9331	1.8632	-0.9354
4	-0.4228	0.0272	0.5673	-0.9354
5	1.637	1.6608	1.5752	-0.9354
6	-0.206	-0.7896	-1.0166	0.4009
7	0.2277	-0.5173	-0.2966	0.4009
8	-1.3985	-0.7896	-1.1606	2.4054
9	-1.0732	-0.5173	-0.4406	-0.2673
10	-0.6396	-0.5173	-0.1526	0.4009

$$D0_{i,k} := \sum_{j=1}^{m} (Z0_{i,j} - Z0_{k,j})^2$$

D0 =

	1	2	3	4	5	6	7	8	9
1	0	5.8399	9.4711	2.745	7.2004	3.2388	1.3262	10.0487	3.5935
2	5.8399	0	21.5054	3.7418	17.5268	1.8598	2.7886	12.6776	1.8268
3	9.4711	21.5054	0	10.4944	0.2041	21.7344	15.0993	38.2939	20.3259
4	2.745	3.7418	10.4944	0	7.9271	5.0086	3.2517	15.7654	2.1819
5	7.2004	17.5268	0.2041	7.9271	0	17.904	12.02	33.8635	16.5995
6	3.2388	1.8598	21.7344	5.0086	17.904	0	0.7805	5.4606	1.6044
7	1.3262	2.7886	15.0993	3.2517	12.02	0.7805	0	7.4826	2.1595
8	10.0487	12.6776	38.2939	15.7654	33.8635	5.4606	7.4826	0	7.8411
9	3.5935	1.8268	20.3259	2.1819	16.5995	1.6044	2.1595	7.8411	0
10	1.9913	3.0167	18.0708	2.6476	14.6983	1.0086	0.7729	5.6838	0.7174

$\text{sort}(D0^{\langle 1 \rangle}) =$

	1
1	0
2	1.3262
3	1.9913
4	2.745
5	3.2388
6	3.5935
7	5.8399
8	7.2004
9	9.4711
10	10.0487

$\text{sort}(D0^{\langle 2 \rangle}) =$

	1
1	0
2	1.8268
3	1.8598
4	2.7886
5	3.0167
6	3.7418
7	5.8399
8	12.6776
9	17.5268
10	21.5054

$\text{sort}(D0^{\langle 3 \rangle}) =$

	1
1	0
2	0.2041
3	9.4711
4	10.4944
5	15.0993
6	18.0708
7	20.3259
8	21.5054
9	21.7344
10	38.2939

Scanned with OKEN Scanner

Scanned with OKEN Scanner

Pasul 1: Se formeaza clusterul format prin reunirea a doi clusteri aflati la cea mai mica distanta(firmele C si E), Acest nou cluster se noteaza CE.

$$CE = \begin{pmatrix} \text{CLUSTER} & \text{CA} & \text{PN} & \text{P} & \text{RP} \\ \text{C} & 45 & 16 & 110 & 3 \\ \text{E} & 43 & 15 & 100 & 3 \end{pmatrix} \qquad CE1 := \begin{pmatrix} 45 & 16 & 110 & 3 \\ 43 & 15 & 100 & 3 \end{pmatrix}$$

$$\text{mean}(CE1^{\langle 1 \rangle}) = 44 \quad \text{stdev}(CE1^{\langle 1 \rangle}) = 1 \quad \text{mean}(CE1^{\langle 2 \rangle}) = 15.5 \quad \text{stdev}(CE1^{\langle 2 \rangle}) = 0.5$$

$$\text{mean}(CE1^{\langle 3 \rangle}) = 105 \quad \text{stdev}(CE1^{\langle 3 \rangle}) = 5 \quad \text{mean}(CE1^{\langle 4 \rangle}) = 3 \quad \text{stdev}(CE1^{\langle 4 \rangle}) = 0$$

$i := 1..9 \qquad j := 1..4 \qquad k := 1..9$

$$C = \begin{pmatrix} \text{CLUSTER1} & \text{CA} & \text{PN} & \text{P} & \text{RP} \\ \text{A} & 30 & 11 & 48 & 5 \\ \text{B} & 26 & 5 & 10 & 3 \\ \text{CE} & 44 & 15.5 & 105 & 3 \\ \text{D} & 24 & 9 & 65 & 5 \\ \text{F} & 26 & 6 & 10 & 5 \\ \text{G} & 30 & 7 & 35 & 5 \\ \text{H} & 15 & 6 & 5 & 8 \\ \text{I} & 18 & 7 & 30 & 4 \\ \text{J} & 22 & 7 & 40 & 5 \end{pmatrix} \qquad C1 := \begin{pmatrix} 30 & 11 & 48 & 5 \\ 26 & 5 & 10 & 3 \\ 44 & 15.5 & 105 & 3 \\ 24 & 9 & 65 & 5 \\ 26 & 6 & 10 & 5 \\ 30 & 7 & 35 & 5 \\ 15 & 6 & 5 & 8 \\ 18 & 7 & 30 & 4 \\ 22 & 7 & 40 & 5 \end{pmatrix}$$

$$\text{mean}(C1^{\langle 1 \rangle}) = 26.1111 \quad \text{stdev}(C1^{\langle 1 \rangle}) = 7.8944$$

$$\text{mean}(C1^{\langle 2 \rangle}) = 8.1667 \quad \text{stdev}(C1^{\langle 2 \rangle}) = 3.0912$$

$$\text{mean}(C1^{\langle 3 \rangle}) = 38.6667 \quad \text{stdev}(C1^{\langle 3 \rangle}) = 29.9035$$

$$\text{mean}(C1^{\langle 4 \rangle}) = 4.7778 \quad \text{stdev}(C1^{\langle 4 \rangle}) = 1.3966$$

$$Z1_{i,1} := \frac{C1_{i,1} - \text{mean}(C1^{\langle 1 \rangle})}{\text{stdev}(C1^{\langle 1 \rangle})} \qquad Z1_{i,2} := \frac{C1_{i,2} - \text{mean}(C1^{\langle 2 \rangle})}{\text{stdev}(C1^{\langle 2 \rangle})}$$

$$Z1_{i,3} := \frac{C1_{i,3} - \text{mean}\left(C1^{\langle 3 \rangle}\right)}{\text{stdev}\left(C1^{\langle 3 \rangle}\right)} \qquad Z1_{i,4} := \frac{C1_{i,4} - \text{mean}\left(C1^{\langle 4 \rangle}\right)}{\text{stdev}\left(C1^{\langle 4 \rangle}\right)}$$

$$Z1 = \begin{pmatrix} 0.4926 & 0.9166 & 0.3121 & 0.1591 \\ -0.0141 & -1.0244 & -0.9586 & -1.2729 \\ 2.266 & 2.3723 & 2.2182 & -1.2729 \\ -0.2674 & 0.2696 & 0.8806 & 0.1591 \\ -0.0141 & -0.7009 & -0.9586 & 0.1591 \\ 0.4926 & -0.3774 & -0.1226 & 0.1591 \\ -1.4075 & -0.7009 & -1.1258 & 2.3071 \\ -1.0275 & -0.3774 & -0.2898 & -0.5569 \\ -0.5208 & -0.3774 & 0.0446 & 0.1591 \end{pmatrix}$$

$$D1_{i,k} := \sum_{j=1}^{4} \left(Z1_{i,j} - Z1_{k,j}\right)^2$$

$$D1 = \begin{pmatrix} 0 & 7.6896 & 10.9482 & 1.3194 & 4.4878 & 1.8634 & 12.9083 & 4.86 & 2.7729 \\ 7.6896 & 0 & 26.8293 & 7.1721 & 2.1553 & 3.4249 & 14.8906 & 2.4055 & 3.7324 \\ 10.9482 & 26.8293 & 0 & 14.6798 & 26.7869 & 18.2363 & 46.9388 & 25.2112 & 22.1027 \\ 1.3194 & 7.1721 & 14.6798 & 0 & 4.3889 & 2.0027 & 10.8814 & 2.8788 & 1.1817 \\ 4.4878 & 2.1553 & 26.7869 & 4.3889 & 0 & 1.0603 & 6.5834 & 2.0916 & 1.3678 \\ 1.8634 & 3.4249 & 18.2363 & 2.0027 & 1.0603 & 0 & 9.3354 & 2.8512 & 1.0549 \\ 12.9083 & 14.8906 & 46.9388 & 10.8814 & 6.5834 & 9.3354 & 0 & 9.1505 & 6.8747 \\ 4.86 & 2.4055 & 25.2112 & 2.8788 & 2.0916 & 2.8512 & 9.1505 & 0 & 0.8812 \\ 2.7729 & 3.7324 & 22.1027 & 1.1817 & 1.3678 & 1.0549 & 6.8747 & 0.8812 & 0 \end{pmatrix}$$

Pasul 2: Se unesc doi clusteri intre care distanta este minima (firmele I si J)

Procedeul continua pana cand se obtine un singur cluster.Se urmareste obtinerea de grupur compacte , omogene cu dispersii cat mai mici.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

CAPITOLUL 13

BALANȚA DE PLĂȚI EXTERNE ANALIZĂ STATISTICĂ

Relațiile economice internaționale ale țării, determină poziția ei în raporturile cu restul lumii, care este reflectată de două documente:

— balanța de plăți externe și

— poziția investițională internațională a țării.

- **BALANȚA DE PLĂȚI EXTERNE** (BPE),

înregistrează operațiunile (tranzacțiile) dintre **rezidenți și străini**, derulate într-o perioadă de referință (de obicei anul calendaristic).

- **POZIȚIA INVESTIȚIONALĂ INTERNAȚIONALĂ** (P.I.I),

Exprimă stocul de *creanțe* (drepturi) și angajamente (obligații) financiare ale unei țări față de alte state la un moment dat, de obicei sfârșitul anului calendaristic.

Variația stocului de creanțe și angajamente se obține în urma tranzacțiilor derulate cu nerezidenții, lucru consemnat și în BPE, fiind o consecință a schimbărilor petrecute în **cursul de schimb** al monedelor, precum și al prețurilor folosite în evaluarea lor sau în urma ajustărilor valorice.

Din aceste documente se deduce și datoria externă a țării acumulată în relațiile cu organisme și / sau instituții financiare internaționale.

13.1. Sistemul conceptual și tipologia BPE

Potrivit **manualului balanței de plăți externe**, elaborat de FMI (Fondul Monetar Internațional), B.P.E este acel document statistic care prezintă tranzacțiile unei economii cu restul lumii pe o perioadă de un an.

Conceptul de **tranzacție internațională** stă la baza elaborării BPE și exprimă **crearea, transformarea, comercializarea**, transferul sau stingerea unei valori economice în relațiile unei economii cu restul lumii.

Tranzacţia presupune:

— schimbul de proprietate asupra bunurilor materiale şi / sau a drepturilor financiare;

— prestarea de servicii;

— disponibilizarea de forţă de muncă sau capital în relaţiile dintre rezidenţii unei economii şi nerezidenţi.

- **Rezidenţii unei economii sunt:**

— familiile şi indivizii care îşi creează în cadrul economiei o gospodărie;

— întreprinderile (societăţii comerciale şi regii autonome), corporaţiile şi cvasi-corporaţiile (sucursale ale investitorilor străini care nu au dobândit personalitate juridică);

— organizaţiile private non-profit;

— administraţia publică centrală (guvernul) şi cea locală.

În rândul nerezidenţilor sunt incluse:

— persoane fizice şi juridice străine, care fiinţează în afara teritoriului economic al ţării;

— co-naţionalii care şi-au mutat centrul de interes economic sau de activitate în străinătate;

— filialele şi sucursalele societăţilor comerciale autohtone stabilite în străinătate;

— turişti străini

— oamenii de afaceri şi funcţionarii diplomatici aflaţi temporar pe teritoriul unei alte ţări decât cea de origine. Se menţionează faptul că accepţiunea economică a rezidenţei este preferată în locul criteriilor legale sau al naţionalităţii.

Centrul de interes economic se află într-o ţară atunci când o persoană fizică sau juridică are plasamente în economia acelei ţări, unde sau de unde s-a angajat sau intenţionează să se angajeze în activităţi şi tranzacţii comerciale sau financiare, pe o perioadă de timp nu mai mică de un an.

Teritoriul economic al ţării corespunde teritoriului geografic administrat de guvernul unei ţări, unde persoanele, bunurile şi capitalul circulă libere.

Scanned with OKEN Scanner

Teritoriul economic al unei țări cuprinde:

— spațiul aerian;

— apele teritoriale și teritorii din apele internaționale asupra cărora țara are drepturi exclusive;

— enclavele teritoriale din restul lumii;

— zonele libere;

— depozitele de graniță și/sau unitățile operate de întreprinderi nerezidente sub controlul vamal al țării respective.

Teritoriul economic al unei țări nu cuprinde enclavele teritoriale ale guvernelor străine sau organizațiilor internaționale aflate în interiorul granițelor politic recunoscute.

Evaluarea tranzacției se efectuează la prețul pieței, acesta fiind înțeles ca suma de bani pentru care operațiunea se încheie în condițiile în care partenerii sunt părți independente, reunite din interese strict economice.

Momentul de înregistrare al tranzacției este momentul (real sau estimat) al schimbului de proprietate. În practica economică s-a convenit ca o aproximare rezonabilă a acestuia, momentul înregistrării tranzacției în evidențele contabile ale societăților bancare la care partenerii au deschise conturile lor.

13.2. **Structura unei B.P.E.**

Totalitatea tranzacțiilor unei țări cu restul lumii sunt repartizate **în două conturi** (în conformitate cu „Manualul BPE al FMI"):

- **Contul curent** sau B.P.E. curente
- Contul de capital și financiar sau balanța mișcărilor de capital

Fiecare cont se detaliază pe subconturi (capitale) posturi și poziții, pentru a evidenția:

— specializarea internațională a economiei țării și

— interesele economice și financiare promovate

Scanned with OKEN Scanner

CONTUL CURENT AL BPE

Cuprinde totalitatea tranzacțiilor **care nu au** ca obiect **capitalul** sau **activele financiare.**

Tranzacțiile curente au ca obiect:

- mărfuri și servicii
- veniturile rezultate din migrația internațională a factorilor sau transferurile internaționale

Contul curent este **divizat** în **trei subconturi** (capitale)

A. BUNURI ȘI SERVICII

Acest capital reunește:

- balanța comercială
- exportul și importul de bunuri materiale (bunuri corporale, tangibile sau vizibile evaluate la prețuri **franco-frontieră vamală** a țării exportatoare, sau prețuri **fob port** de îmbarcare (fob. — free on board = liber la bordul navei) și
- balanța serviciilor, care cuprinde **încasări** și **plăți** rezultate din prestarea de servicii între rezidenți și nerezidenți.

B. **CAPITOLUL VENITURI** sau balanța veniturilor cuprinde încasări (+) și plăți (–) cauzate de faptul că factorii de producție trec frontiera țării pentru a se valorifica, iar veniturile rezultate din acest procedeu sunt total sau parțial repatriate.

Aici se înregistrează:

- **veniturile din muncă;**
- **veniturile din investiții directe și de portofoliu;**
- dobânzi (venituri ale capitalului împrumutat)

C. TRANSFERURILE CURENTE

Se referă la intrările și / sau ieșirile de

- **resurse reale** (bunuri și / sau servicii) sau
- **financiare** (cotizații, burse, premii, ajutoare financiare nerambursabile, etc.) fără compensare din partea beneficiarului.

Astfel de transferuri sunt făcute de:

Scanned with OKEN Scanner

- **instituţiile administraţiei publice** (contribuţii la bugetele administrative ale organizaţiilor internaţionale, cheltuieli pentru întreţinerea reprezentanţelor diplomatice, consulare, culturale, militare, etc., burse de stat, premii internaţionale ajutoare nerambursabile etc. acordate străinătăţii şi respectiv primite din străinătate), fie de persoane fizice sau organizaţii private, în conformitate cu legislaţia în vigoare şi potrivit intereselor pe care le promovează (chirii, moşteniri, premii, burse, donaţii, ajutoare nerambursabile etc., operate între rezidenţi şi nerezidenţi).

CONTUL DE CAPITAL ŞI FINANCIAR al BPE

Evidenţiază mişcările de capital grupate în două (subconturi distincte în funcţie de natura activelor tranzacţionate).

- **CONTUL DE CAPITAL**

Cuprinde toate operaţiunile care au ca obiect **achiziţionarea** şi **vânzarea** de **active nefinanciare** care nu sunt rezultatul activităţii umane (cum sunt terenurile sau bogăţiile subsolului), precum şi **active intangibile** (brevete, mărci, copyright etc.).

- B **CONTUL FINANCIAR**

Reuneşte operaţiunile care au ca obiect **active financiare**, inclusiv crearea şi/sau lichidarea de creanţe în relaţia cu restul lumii. **Aceste operaţiuni** sunt structurate pe patru tipuri de tranzacţii.

- B_1 **Investiţii directe,** care presupun plasamente financiare într-o întreprindere situată în afara teritoriului economic al ţării investitorului, în condiţiile unei atitudini active în luarea deciziilor în întreprindere.

NOTĂ: Fără a fi o recomandare FMI, limita inferioară a unui asemenea plasament trebuie să fie de cel puţin 10% din capitalul social al firmei pentru a putea fi considerată ca investiţie directă. Acest flux cuprinde atât plasamentele iniţiale, cât şi câştigurile reinvestite.

- B_2 **Investiţii de portofoliu**

Sunt plasamente financiare într-o societate comercială sau bancară situată în afara teritoriului economic al ţării investitorului cu scopul de a asigura o valorificare mai avantajoasă a capitalului pe piaţa externă decât se poate obţine pe propria piaţă financiară.

Scanned with OKEN Scanner

Obiectul acestor tranzacţii îl constituie **acţiunile şi obligaţiunile** negociate pe piaţa bursieră sau extrabursieră.

- B_3 **Alte investiţii**

Această poziţie se referă la împrumuturile de la FMI şi folosirea acestora, depozite bancare, credite comerciale acordate / primite în relaţiile dintre rezidenţi şi nerezidenţi.

- B_4 **Activele de rezervă**

Se află sub controlul şi la dispoziţia autorităţii monetare din fiecare ţară (Banca Naţională a României — BNR).

Activele de rezervă sunt constituite din:

- aur monetar
- drepturi speciale de tragere
- poziţia de rezervă la FMI
- disponibilităţi în valute străine
- alte creanţe.

Poziţia din BPE reflectă doar mişcarea acestor active ca urmare a activităţii autorităţii monetare centrale pentru echilibrarea BPE, pentru investiţii pe piaţa valutară în vederea influenţării cursului de schimb al monedei naţionale etc.

NOTĂ: Poziţiile **conturi în tranzit** şi conturi de cliring / barter conţin sume a căror provenienţă sau destinaţie este în curs de stabilire şi respectiv, **tranzacţii internaţionale** care au ca obiect comercializarea de bunuri materiale şi / sau servicii fără a folosi mijloace de plată (clirinţuri, operaţiuni de barter etc.).

- POZIŢIA de **ERORI ŞI OMISIUNI** apare în fiecare BPE cu o anumită încărcare fie pe **debit** fie pe credit ca urmare a faptului că tranzacţiile comerciale şi financiare derulate de băncile comerciale îmbracă de cele mai multe ori o formă complexă ce nu permite întotdeauna identificarea uşoară şi imediată a structurii analitice propuse de FMI prin Manualul B.P.E. Alteori tranzacţii derulate în chiar ziua prezentării la BNR a rapoartelor lunare de către băncile comerciale rămân să fie adăugate abia la sfârşitul anului.

Omisiunile şi inadvertenţele tind să se compenseze. Dacă însă diferenţele, prezentate la omisiuni şi erori, sunt mari şi persistă an de an,

Scanned with OKEN Scanner

credibilitatea informațiilor cuprinse în BPE tinde să fie pusă sub semnul întrebării, iar analiza diferitelor poziții ale balanței rămâne la nivelul unei tentative de caracterizare. În asemenea condiții, BPE nu mai reprezintă un instrument de analiză macroeconomică.

13.3. Echilibrarea balanței globale de P.E.

- **SOLDUL** unei balanțe se obține ca diferență (sumă netă) între **încasările (*creditul*)** și **plățile (*debitul*)** aferente tranzacțiilor înscrise în acea balanță.

Rareori se întâmplă ca balanța să fie echilibrată, adică să prezinte sold zero, adică încasările să fie egale cu plățile. De obicei se constată o de balanță sau un dezechilibru, fie într-un sens fie în celălalt. Astfel dacă încasările sunt mai mari decât plățile, soldul este pozitiv sau excedentar (balanța activă sau excedentară). Dacă încasările sunt inferioare plăților, atunci soldul este negativ sau deficitar (balanță pasivă sau deficitară)

Ca urmare a dublei înregistrări a tranzacțiilor cu restul lunii, BPE a unei țări este în mod necesar echilibrată, în sensul că la nivel global credit = debit sau încasări = plăți, soldul este nul.

Conceptele de dezechilibru, de balanță, activ / pasiv, excedent / deficit sunt aplicabile în cazul unei balanțe globale de plăți externe numai conturilor, capitalelor sau pozițiilor din balanță, nu şi întregii B.P.E. În acest context B.P.E. este împărțită în două conturi:

- **Contul curent** aflat „peste linie“ și
- **Contul de capital şi financiar** situat „sub linie“, deoarece dezechilibrul constatat deasupra liniei în contul curent se reglează prin balanța de sub linie.

În măsura în care soldurile posturilor (componentelor) **contului curent** nu se compensează, diferența pozitivă sau negativă este absorbită prin posturile (pozițiile) aflate sub linie, iar în ultimă instanță ea grevează asupra activelor rezervă ale statului.

Scanned with OKEN Scanner

INFLUENȚA EXCEDENTULUI CONTULUI CURENT AL BPE

Excedentul contului curent al BPE poate determina „sub linie" fie:

- un export de capital sub formă de investiție (**directă** sau de portofoliu) în străinătate sau un împrumut acordat nerezidenților,
- reducere a propriei datorii externe față de străinătate prin achitarea (restituirea) unor tranșe din împrumuturile contractate
- creșterea rezervelor oficiale ale țării.

INFLUENȚA DEFICITULUI CONTULUI CURENT AL BPE

Un deficit al contului curent al BPE generează „**sub linie**", adică în contul de capital şi financiar una din următoarele operațiuni:

- un import de capital (investiție directă, investiție de portofoliu sau împrumut contractat);
- amânare a rambursării datoriei externe;
- Cheltuirea unei părți din rezervele oficiale pentru a achita plățile scadente în perioada de referință.

Evident echilibrarea BPE poate determina folosirea concomitentă a două sau chiar a celor trei posibilități evocate pentru fiecare situație (excedent sau deficit).

La nivel regional şi/sau bilateral, balanțele de plăți externe sunt, de cele mai multe ori, dezechilibrate.

Banca Națională a României în calitatea ei de autoritate monetară centrală are misiunea de a supraveghea aceste dezechilibre „parțiale" şi de a sugera Guvernului măsuri de politică financiară, comercială şi economică menite să corecteze în timp aceste debalanțe, astfel încât ele să nu afecteze grav echilibrul global al schimburilor țării cu restul lumii.

13.4. **Analiza statistică a BPE**

Dezechilibrul consemnat în cadrul fiecărui cont se analizează prin:

a) **mărimea absolută** a dezechilibrului, exprimat **prin soldul fiecărei balanțe, respectiv soldul fiecărei componente**;

Scanned with OKEN Scanner

b) **mărimea relativă** a dezechilibrului, pentru a cunoaşte gravitatea sau amploarea acestei stări;
c) **evoluţia în timp** a dezechilibrului;
d) **structura dezechilibrului**, ceea ce permite explicarea originilor sale;
e) comparaţii între ţări cu grad similar de dezvoltare sau aflate în aceeaşi grupare de state.

Indicatori pentru analiza contului curent al BPE

Notăm:
— încasările generate de:
xj — exportul de mărfuri sau de servicii (**care exprimă toate înregistrările din creditul BPE**)
mj — plăţile pentru import (**care asimilează înregistrările din debitul BPE**)
Sj — soldul unei anumite poziţii j ($j = \overline{1,k}$) componente, poziţii sau posturi ale balanţei,
şi obţinem:

$$Sj = xj - mj \gtrless 0 \qquad (13.1)$$

iar pentru **întregul cont curent**:

$$S = X - M \gtrless 0 \qquad (13.2)$$

$$\text{unde } S = \sum_{j=1}^{K} Sj = \sum_{j=1}^{K} (xj - mj) = \sum_{j=1}^{K} xj - \sum_{j=1}^{K} mj = X - M$$

Mărimea absolută a soldului contului curent sau a soldului uneia din componentele sale constituie o informaţie utilă pentru a putea continua analiza, dar ea exprimă diferenţa între încasările şi plăţile aferente postului.

Deficitul **contului curent** al BPE reprezintă o constatare care nu admite emiterea vreunei aprecieri calitative de felul: deficitul e mare sau

Scanned with OKEN Scanner

e mic, e grav, nu e neglijabil. Abia compararea soldului cu alți indicatori conduce la obținerea de mărimi derivate (relative) ale soldului, care exprimă cât de grav este dezechilibrul.

O astfel de mărime derivată arată procentual cât reprezintă deficitul (excedentul) **contului curent față de PIB**.

Se apreciază că un dezechilibru de ± 2% până la ± 5% din PIB firesc, cu condiția ca debalanța să nu se permanentizeze. Când se trece de ± 5% față de PIB, dezechilibrul devine important şi reclamă măsuri concrete ale Guvernului şi autorității monetare (BNR) pentru a redresa balanța într-un termen destul de scurt (rezonabil).

Dacă dezechilibrul depăşeşte ±10% față de PIB, el este semnul unor disfuncții grave în economie, măsurile de natură fiscală sau monetară nemaifiind suficiente. Ele trebuie însoțite **de energice politici sectoriale**, care pot corecta dezechilibrul după cel puțin 2–3 ani de aplicare consecventă

- **Indicatorul** MRS_1 (mărimea relativă a soldului) **contului curent** față de PIB:

$$MRS_1 = \frac{S}{PIB} \cdot 100 \qquad (13.3)$$

În literatura de specialitate acest indicator este frecvent comentat nefavorabil, ca urmare a faptului că doar numărătorul este absolut fiabil (evidența BPE se ține în dolari SUA), pe când muncitorul este determinat în moneda națională, iar procedeele de conversie a expresiei valorice în USD sunt discutabile.

- **Indicatorul** $\mathbf{MRS_2}$ (mărimea relativă a soldului) contului curent față de totalul tranzacțiilor curente (X + M):

$$MRS_2 = \frac{S}{X + M} \cdot 100 \qquad (13.4)$$

De data aceasta debalanța este apreciată ca fiind în limite acceptabile dacă rămâne inferioară nivelului de ±10% față de totalul tranzacțiilor curente. Un avantaj suplimentar al acestei mărimi relative constă în faptul că debalanța poate fi calculată nu numai la nivelul întregii BPE curente, ci şi pe capitole, poziții, posturi, etc., din folosirea relației:

Scanned with OKEN Scanner

$$(MRS_2)j = \frac{Sj}{xj + mj} \cdot 100 \qquad (13.5)$$

- **Gradul de acoperire** (G.A.) a **plăților prin încasări** este un indicativ relativ care exprimă gravitatea debalanței prin abaterea față de poziția de echilibru (100%). Gradul de acoperire este **expresia procentuală** a soldului balanței contului curent:

$$G.A = \frac{X}{M} \cdot 100 \gtrless 100\% \qquad (13.6)$$

iar pentru fiecare componentă:

$$g.a\ j = \frac{xj}{mj} \cdot 100 \gtrless 100\% \qquad (13.7)$$

Spre deosebire de soldul balanței, care poate avea semnul + nu –, gradul de acoperire a plăților prin încasări prezintă avantajul evitării semnului:

- unui sold excedentar (+) îi corespunde un grad de acoperire GA >100%, în timp ce un sold deficitar (–) poate fi exprimat printr-un grad de acoperire GA < 100%.

Mărimea absolută și relativă a soldului înregistrat de componentele contului curent al BPE a României în 1993–1995

Indicatorul	U.M.	1993	1994	1995
I. Mărimea absolută a soldului				
cont curent al BPE	md. USD	1174	428	1336
– balanța comercială	- " -	1128	411	1231
– balanța serviciilor	- " -	115	171	233
– balanța veniturilor	- " -	145	129	241
– balanța transferurilor	- " -	+214	+283	+369

Scanned with OKEN Scanner

II. Mărimea relativă a soldului contului curent al BPE față de:				
PIB	procente	4,5	1,5	3,8
totalul tranzacțiilor curente	- ” -	8,9	2,6	6,5
III. Mărimea relativă a soldului înregistrat de componentele contului curent *față de total tranzacții*:				
– balanța comercială	procente	10,3	3,2	7,6
– balanța serviciilor	- ” -	7,6	7,6	7,2
– balanța veniturilor	- ” -	53,5	35,7	59,8
– balanța transferurilor	- ” -	+86,3	+69,9	+62,6
IV. Gradul de acoperire a plăților prin încasări				
cont curent	procente	86,6	94,7	87,8
– balanța comercială	- ” -	81,3	93,7	85,9
– balanța serviciilor	- ” -	87,4	85,9	86,5
– balanța veniturilor	- ” -	30,3	47,3	25,2
– balanța transferurilor	- ” -	1358,8	563,9	435,5

Sursa: calcule bazate pe rapoartele anuale ale BNR cu privire la balanța de plăți externe din 1993–1995.

Scanned with OKEN Scanner

PROGRAM PENTRU ANALIZA STATISTICA A BALANTEI DE PLATI EXTERNE A ROMANIEI

Notatii:

x1=CONTUL CURENT - CREDIT
x11=Bunuri si servicii-credit
x12=Venituri-credit
x13=Transferuri curente-credit

x2=CONTUL CURENT - DEBIT
x21=Bunuri si servicii-debit
x22=Venituri-debit
x23=Transferuri curente-del

y1=CONTUL DE CAPITAL SI FINANCIAR-CREDIT
y11=Contul de capital-credit
y12=Contul financiar-credit

y2=CONTUL DE CAPITAL SI FINANCIAR-DEBIT
y21=Contul de capital-debit
y22=Contul financiar-debit

EO1=ERORI SI OMISIUNI-CREDIT

OE2=ERORI SI OMISIUNI-DEBIT

BALANTA DE PLATI EXTERNE (BPE) a ROMANIEI in perioada 1993-1996 s caracterizeaza prin urmatoarele date (prezentate ca vectori cu indicii i=1,2,3,4 corespunzatori anilor 1993,1994,1995 si respectiv 1996) exprimate in mil.$

ORIGIN≡ 1

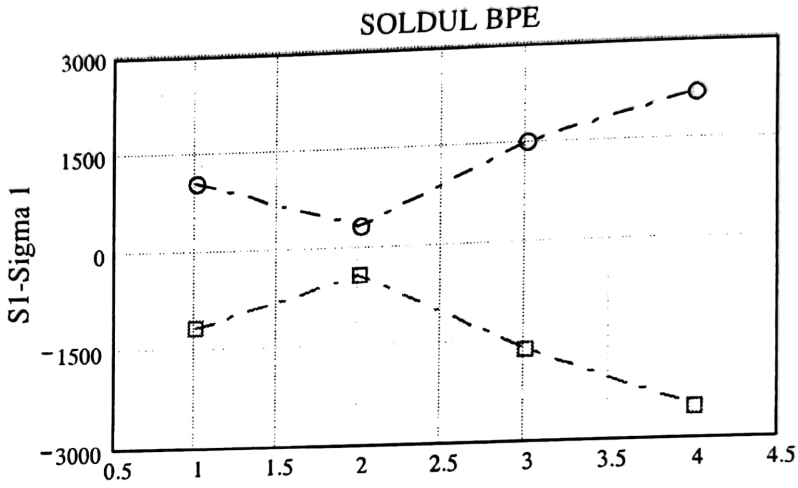
$$x1:=\begin{pmatrix}5985\\7655\\9964\\10402\end{pmatrix}\quad x11:=\begin{pmatrix}5691\\7195\\9404\\9648\end{pmatrix}\quad x12:=\begin{pmatrix}63\\116\\81\\79\end{pmatrix}\quad x13:=\begin{pmatrix}231\\344\\479\\675\end{pmatrix}\quad OE1:=\begin{pmatrix}101\\92\\123\\0\end{pmatrix}$$

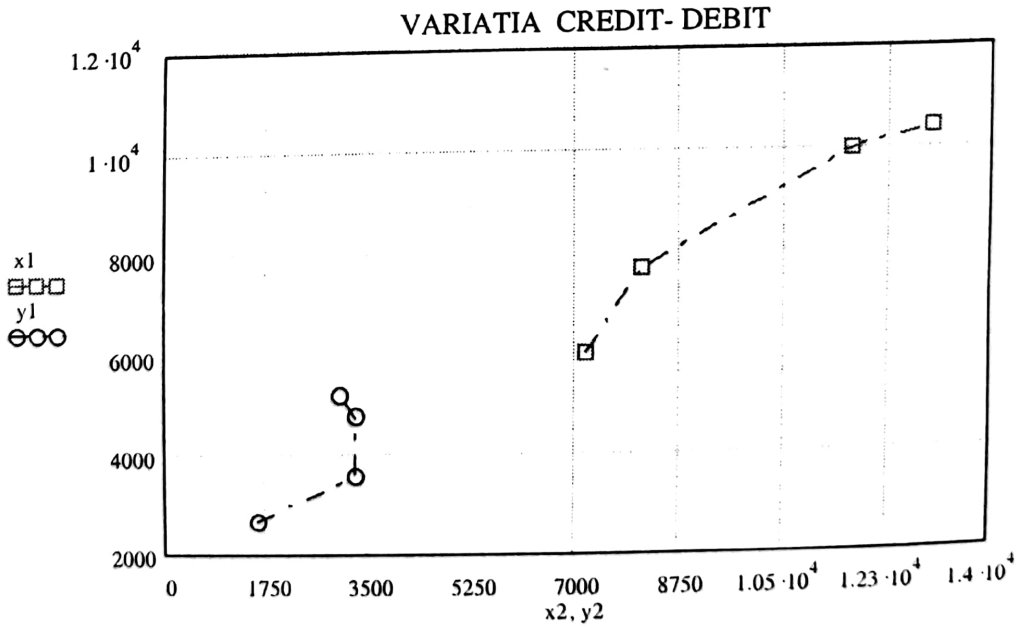
$$x2:=\begin{pmatrix}7155\\8083\\11603\\12973\end{pmatrix}\quad x21:=\begin{pmatrix}6930\\7777\\11171\\12503\end{pmatrix}\quad x22:=\begin{pmatrix}208\\245\\322\\388\end{pmatrix}\quad x23:=\begin{pmatrix}17\\61\\110\\82\end{pmatrix}\quad OE2:=\begin{pmatrix}0\\0\\0\\357\end{pmatrix}$$

$$y1:=\begin{pmatrix}2651\\3604\\4781\\5206\end{pmatrix}\quad y11:=\begin{pmatrix}8\\12\\242\\152\end{pmatrix}\quad y12:=\begin{pmatrix}2643\\3592\\4539\\5054\end{pmatrix}\quad y2:=\begin{pmatrix}1582\\3270\\3265\\2992\end{pmatrix}\quad y21:=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}\quad y22:=\begin{pmatrix}1582\\3270\\3265\\2992\end{pmatrix}$$

Scanned with OKEN Scanner

$$\Sigma 1 = \begin{pmatrix} 1.069\times 10^3 \\ 334 \\ 1.516\times 10^3 \\ 2.214\times 10^3 \end{pmatrix} \qquad \Sigma 11 = \begin{pmatrix} 8 \\ 12 \\ 242 \\ 152 \end{pmatrix} \qquad \Sigma 12 = \begin{pmatrix} 1.061\times 10^3 \\ 322 \\ 1.274\times 10^3 \\ 2.062\times 10^3 \end{pmatrix}$$





Scanned with OKEN Scanner

ANEXA 1

UTILIZAREA MATHCAD IN STATISTICA

Produsul MATHCAD, elaborat de firma Mathsoft Inc. Cambridge, MA, SUA, utilizabil pe microcalculatoare compatibile IBM are scopul de a asista pe cercetători, din diferite domenii, care în activitatea lor au nevoie de rezolvarea unor probleme de matematică sofisticate.

În prezentul capitol vor fi abordate unele aplicații statistice cu ajutorul mathCAD-ului.

Funcții mathcad

În prezentarea funcțiilor se notează cu :

z –un număr complex
x,y – numere reale
n – reprezintă un număr natural
v – reprezintă un vector
M – reprezintă o matrice

	APELAREA FUNCȚIEI	SEMNIFICAȚIA.
1	sin(z)	sinz
2	cos(z)	cosz
3	tan(z)	tgz
4	asin(z)	arcsinz
5	acos(z)	arccosz
6	atan(z)	arctgz
7	angle(x,y)	argumentul lui z=x+iy
8	sinh(z)	sinhz
9	cosh(z)	coshz
10	tanh(z)	tghz(thz)
11	asinh(z)	arcshz

Scanned with OKEN Scanner

12	acosh(z)	arcchz
13	atanh(z)	arcthz
14	exp(z)	expz(e la z)
15	ln(z)	lnz
16	log(z)	lgz
17	Γ(z)	Funcția Γ a lui Euler
18	erf(x)	funcția erorilor
19	cnorm(x)	funcția de repartiție normală
20	floor(x)	partea întreagă a lui x
21	ceil(x)	floor(x)+1
22	rnd(x)	Număr aleator cuprins între 0 şi x
23	Re(z)	Rez
24	Im(z)	Imz
25	arg(z)	argz
26	mod(x,y)	Restul împărțirii întregi a lui x la y
27	length(v)	Numărul elementelor vectorului v
28	last(v)	Indicelei ultimului element al vectorului v
29	max(v)	Cel mai mare element al vectorului v
30	rows(M)	Numărul liniilor matricei M
31	cols(M)	Numărul coloanelor matricei M
32	tr(M)	Urma matricei M
33	augment(M1,M2)	Matricea obținută prin alăturarea matricelor M1,M2 cu acelaşi număr de linii
34	identity(n)	Matricea unitate de ordin n
35	root(expr,var)	Returnează valoarea variabilei var care anulează expresia expr.
36	if(cond,z1,z2)	Returnează z1 dacă condiția este adevărata, altfel se returnează z2
37	until(x1,x2)	Returnează x2 cât timp x1>0.Dacă x1<0 atunci calculul se opreşte
38	sort(v)	Ordonează crescător elementele vectorului v
39	csort(M,n)	Aranjează liniile matricei M astfel încât elementele coloanei n să fie în ordine crescătoare
40	rsort(M,n)	Aranjează coloanele matricei M astfel încât elementele liniei n să fie în ordine crescătore
41	reverse(v),reverse(M)	Scrie elementele vectorului v în ordine inversă (de la drapta la stânga) şi formează o matrice scriind liniile lui M de jos în sus

Scanned with OKEN Scanner

42	mean(v)	Valoarea medie a elementelor vectorului v
43	var(v)	Dispersia elementelor lui v
44	stdev(v)	Abaterea medie pătratica
45	corr(vx,vy)	Coeficientul de corelație al lui Pearson aplicat vectorilor de date vx,vy
46	slope(vx,vy)	Panta ecuației regresiei liniare
47	intercept(vx,vy)	Ordonata la origine a ecuației regresiei liniare
48	cspline(vx,vy)	Calculează coeficienții funcției spline cubice de interpolare determinată de vectorii de date vx,vy .
49	interp(vs,vx,vy,x)	Returnează valoarea în x a funcției spline cubice ;vs este vectorul cu coeficienții funcției spline calculat în cspline(vx,vy)
50	δ(n1,n2)	Funcția lui Kronnecker

Utilizarea calculului matricial în mathcad

Generarea matricilor :

ORIGIN≡1 $\quad i := 1..3 \quad j := 1..5 \quad A_{i,j} := rnd(8) \quad B_{i,j} := rnd(10)$

$$A = \begin{bmatrix} 0.01 & 1.547 & 4.68 & 2.802 & 6.583 \\ 1.393 & 5.684 & 2.432 & 0.731 & 1.179 \\ 7.908 & 0.953 & 0.071 & 4.253 & 4.814 \end{bmatrix} \qquad B = \begin{bmatrix} 1.662 & 4.508 & 0.571 & 7.833 & 5.199 \\ 8.76 & 9.559 & 5.393 & 4.621 & 8.622 \\ 7.797 & 9.968 & 6.115 & 2.662 & 8.401 \end{bmatrix}$$

— prin introducerea elementelor matricei în locurile indicate în formatul dat de meniul Insert al Mathcad-ului.

$$C = \begin{bmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{bmatrix} \qquad C := \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ -3 & 5 & -7 \end{bmatrix} \qquad D = \begin{bmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{bmatrix} \qquad D := \begin{bmatrix} 9 & 0 \\ 1 & 4 \\ 5 & -8 \\ 3 & 7 \end{bmatrix}$$

Scanned with OKEN Scanner

$$\text{identity}(2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{identity}(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad O_3 = \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{pmatrix} \qquad O_3 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

— prin alipirea a doua matrice de acelaşi tip ,utilizând funcţia „augment".

$$N := \text{augment}(C, \text{identity}(3)) \qquad N = \begin{pmatrix} 1 & 3 & 5 & 1 & 0 & 0 \\ 2 & 4 & 6 & 0 & 1 & 0 \\ -3 & 5 & -7 & 0 & 0 & 1 \end{pmatrix}$$

Vectorii pot fi definiţi cu ajutorul matricelor (matrice coloană) sau pot fi extraşi din matrice existente :

$$a := \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad C^{\langle 1 \rangle} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \qquad C^{\langle 2 \rangle} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \qquad C^{\langle 3 \rangle} = \begin{bmatrix} 5 \\ 6 \\ -7 \end{bmatrix} \qquad D^{\langle 1 \rangle} = \begin{bmatrix} 9 \\ 1 \\ 5 \\ 3 \end{bmatrix} \qquad D^{\langle 2 \rangle} = \begin{bmatrix} 0 \\ 4 \\ -8 \\ 7 \end{bmatrix}$$

Operaţii cu matrice :

1) Adunarea matricelor :-se pot aduna numai matrice de acelaşi tip

$$A + B = \begin{bmatrix} 7.816 & 10.631 & 10.467 & 6.05 & 6.71 \\ 11.99 & 12.567 & 9.986 & 16.016 & 5.496 \\ 6.915 & 9.637 & 5.052 & 16.21 & 11.916 \end{bmatrix}$$

2) Înmulţirea cu scalar :

$$\alpha := 3 \qquad \alpha \cdot C = \begin{bmatrix} 3 & 9 & 15 \\ 6 & 12 & 18 \\ -9 & 15 & -21 \end{bmatrix} \qquad C \cdot \alpha = \begin{bmatrix} 3 & 9 & 15 \\ 6 & 12 & 18 \\ -9 & 15 & -21 \end{bmatrix}$$

Scanned with OKEN Scanner

3) Înmulţirea matricelor :

două matrice se pot înmulţi dacă numărul coloanelor matricei deînmulţit este egal cu numărul liniilor matricei înmulţitor. Matricea produs va avea numărul liniilor egal cu numărul liniilor matricei deînmulţit şi numărul de coloane egal cu numărul coloanelor matricei înmulţitor.

$$C^2 = C \cdot C \quad C^3 = C^2 \cdot C \quad C^n = C^{n-1} \cdot C \quad C^2 = \begin{bmatrix} -8 & 40 & -12 \\ -8 & 52 & -8 \\ 28 & -24 & 64 \end{bmatrix} \quad C^3 = \begin{bmatrix} 108 & 76 & 284 \\ 120 & 144 & 328 \\ -212 & 308 & -452 \end{bmatrix}$$

Transpusa unei matrice :

$$C = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ -3 & 5 & -7 \end{bmatrix} \quad C^T = \begin{bmatrix} 1 & 2 & -3 \\ 3 & 4 & 5 \\ 5 & 6 & -7 \end{bmatrix}$$

Definiţie : Dacă este îndeplinită condiţia $X = X^T$, X se numeşte matrice simetrică.

Inversa unei matrice.

O matrice admite inversa dacă este pătratică şi nesingulară (determinantul matricei este nenul) :

$$C = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ -3 & 5 & -7 \end{bmatrix} \quad \det(C) = |C| \quad |C| = 40 \quad C^{-1} = \begin{bmatrix} -1.45 & 1.15 & -0.05 \\ -0.1 & 0.2 & 0.1 \\ 0.55 & -0.35 & -0.05 \end{bmatrix}$$

$$C \cdot C^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad C^{-1} \cdot C = \begin{bmatrix} 1 & 1.069 \cdot 10^{-15} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Scanned with OKEN Scanner

Suma pătratelor elementelor unui vector :

$$v := \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \quad v \cdot v = 29 \quad \text{sau} \quad v^T \cdot v = (29) \quad \text{sau inca} \quad \sum_{i=1}^{3} (v_i)^2 = 29$$

Rădăcina pătrată din elementele unei matrice :

$$A1 := \begin{bmatrix} 1 & 4 & 9 \\ 16 & 25 & 36 \\ 10 & 49 & 81 \end{bmatrix} \qquad \overrightarrow{\sqrt{A1}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 3.162 & 7 & 9 \end{bmatrix}$$

Inversarea elementelor unei matrice :

$$A1 = \begin{bmatrix} 1 & 4 & 9 \\ 16 & 25 & 36 \\ 10 & 49 & 81 \end{bmatrix} \qquad \overrightarrow{\frac{1}{A1}} = \begin{bmatrix} 1 & 0.25 & 0.111 \\ 0.062 & 0.04 & 0.028 \\ 0.1 & 0.02 & 0.012 \end{bmatrix}$$

Rezolvarea sistemelor liniare de tip CRAMER cu ajutorul matricei coeficienților :

$$B1 \cdot X = b \qquad X = \begin{bmatrix} x1 \\ x2 \\ x3 \end{bmatrix} \qquad B1 := \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \qquad b := \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} \qquad B1^{-1} = \begin{bmatrix} 0.75 & -0.25 & -0.25 \\ -0.25 & 0.75 & -0.25 \\ -0.25 & -0.25 & 0.75 \end{bmatrix}$$

$$B1 \cdot B1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad X := B1^{-1} \cdot b \qquad X = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Scanned with OKEN Scanner

ANEXA 2

SPSS

SPSS — sistem de programe de calculator cu aplicații în STATISTICĂ, efectuează paşii necesari analizei listelor de date reprezentând *tabele* şi statistici descriptive, precum şi analiza statistică complexă —

— oferă opțiuni şi caracteristici pentru înțelegerea şi comunicarea rezultatelor sub

Window'95 şi Windows NT — se realizează procesul de analiză a datelor.

OUTPUT NAVIGATOR — permite selectarea rapidă a *tabelelor* şi *graficelor* în ordinea dorită —

Opțiunile SPSS sunt :

— SPSS Professional Statistics
— SPSS Advanced Statistics
— SPSS Tables™
— SPSS Trends™
— SPSS Categories®
— SPSS Chaid™
— SPSS Exact Tests™
— Neural Connection™
— Mapinfo®
— all CLEAR™ III

Compatibilitatea

Sistemul SPSS Base 7.5. — operează pe un sistem — în Windows 95 sau Windows NT

Generalități privind analiza datelor în SPSS.

— sensul *analizei datelor* — variază — de la persoană la persoană
— de la o aplicație la alta

— SPSS este un program cuprinzător, cu multiple *proceduri* ca: liste de date, tabele, statistici descriptive şi analize statistice complexe

Scanned with OKEN Scanner

— grafice pentru selectarea datelor, înțelegerea și interpretarea analizelor și comunicarea rezultatelor care permit statisticienilor să interpreteze rezultatele.

Sistemul SPSS — are instalat și *un ghid* care ilustrează modalitățile de navigare și folosire a diferitelor ferestre și elemente ale SPSS (*Ghidul de utilizare SPSS Base pemtru. Windows).*

Datele SPSS permit să se citească mai multe feluri de **fișiere de date** sau **să se introducă** date direct în **Editorul de date al SPSS.** (Datele sunt prezentate în: **rânduri** (*cazurile*) și **coloane** (variabile)).

Un caz conține informații pentru **o unitate de analiză**, precum o persoană, un animal, o companie, sau un reactor.

Variabilele sunt *informațiile adunate pentru fiecare caz*, cum ar fi înălțimea, greutatea, profitul sau consumul de combustibil. Se pot atribui etichete descriptive numelor variabilelor și valorilor individuale ale fiecărei variabile (ex. 1 = Asia, 2 = Europa ș.a.m.d) sau se afișează în rezultate etichetele sau numele.

Mărimi din Editorul de date.

Pentru informații privind etichetarea, valorile absente, formatele datelor și alte elemente de definire a datelor se folosește *Ghidul de utilizare.*

Rezultatul analizei apare în programul SPSS. Output Navigator.

Tabele pivot se echitează în diferite prețuri pentru afișarea rezultatelor într-o formă care evidențiază relații importante pentru analiză.

Utilizarea programului SPSS

- **Gruparea și reprezentarea grafică a datelor**

Programul de analize statistice SPSS conține proceduri care permit analiza univariată a caracteristicilor populațiilor supuse observării.

Opțiunea generală „Statistics" din meniul principal conține: Summarize cu:

— Frequencies
— Descriptives
— Explore

Pe baza datelor definite prin variabilele var. 1, var. 2,... și apelarea opțiunii Frequencies, se obține tabelul cu distribuția de frecvențe relative

Scanned with OKEN Scanner

şi absolute ale caracteristicii din var. 1 sau altă variabilă, precum şi valorile anumitor indicatori de centralitate şi dispersie. Tot prin această opțiune se pot obține reprezentări grafice ale distribuțiilor. Se apelează Graphs → Inter–active → Bar

Pie → Simple
Clustered
Platted

Se observă că programul afişează sintaxa rulată, apoi un tabel ce conține date despre variabilă: eticheta variabilei, numărul de cazuri valide, respectiv numărul de cazuri care lipsesc. De menționat că există o mulțime de posibilități de expresie grafică cu ajutorul meniului Graphs.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

ANEXA 3

EROARI

Valorile datelor într-un calcul numeric sunt rareori exacte, fiind obținute în urma unor experiențe sau estimări, iar calculul numeric în sine introduce erori de diferite tipuri.

Clasificarea erorilor

- În funcție de formula de calcul a erorii avem:
  - — erori (absolute) și
  - — erori relative (în valoare absolută)
- din punctul de vedere al sursei erorii (cauza erorii) deosebim:
  - — erori inerente
  - — erori de calcul
  - — erori de metodă

Erori absolute și erori relative

Fie x valoarea adevărată a unei date numerice și $\tilde{x}$ valoarea aproximativă a acestei date (obținută din măsurări, observații sau calcule)

Definiții

1. Se numește *eroare conținută* în $\tilde{x}$ și se notează $e_{\tilde{x}}$ diferența $x - \tilde{x}$

$$e_{\tilde{x}} = x - \tilde{x} \qquad (1)$$

2. Se numește modulul *eroării absolute* conținută în $\tilde{x}$ modulul erorii:

$$|e_{\tilde{x}}| = |x - \tilde{x}| \qquad (2)$$

3. Se numește *eroare relativă* conținută în $\tilde{x}$, notată $r_{\tilde{x}}$

$$r_{\tilde{x}} = \frac{e_{\tilde{x}}}{\tilde{x}} \qquad (3)$$

4. Se numeşte *eroare relativă* în valoare absolută conţinută în $\tilde{x}$, modulul erorii relative:

$$|r_{\tilde{x}}| = \left|\frac{e_{\tilde{x}}}{\tilde{x}}\right| = \frac{|e_{\tilde{x}}|}{|\tilde{x}|} \qquad (4)$$

Observaţii

Cantitatea de informaţie furnizată de eroarea relativă este mai mare decât cea furnizată de eroare.

Exemplu: — dacă $x = 9$ şi $\tilde{x} = 10$

$$e_{\tilde{x}} = x - \tilde{x} = 9 - 10 = -1$$

— dacă $y = 999$ şi $\tilde{y} = 1000$

$$e_{\tilde{y}} = 999 - 1000 = -1$$

Se observă că $e_{\tilde{x}} = e_{\tilde{y}}$ pe când:

$$r_{\tilde{x}} = -\frac{1}{10} = -0{,}1; \quad r_{\tilde{y}} = \frac{-1}{1000} = -0{,}001$$

Uneori se poate estima o margine superioară a erorii absolute sau a erorii relative în valoare absolute.

Erori inerente, erori de calcul şi erori de metodă

Fiecare dintre aceste erori (*inerente, de calcul, de metodă*) se pot exprima cu formule ce conţin eroarea absolută sau eroarea relativă.

- **Eroarea totală** conţinută în rezultatul unui calcul numeric se compune din erori inerente, erori de calcul şi erori de metodă.

Eroarea inerentă

Este cauzată de greşeli în observaţie, greşeli în măsurare, greşeli în înregistrare, transmisie, conversie sau greşeli de prepocesare sau greşeli care apar din necesitatea reprezentării unei valori printr-un număr finit de cifre.

Exemple : $\frac{1}{3} = 0{,}33$ sau $\frac{1}{3} = 0{,}333333$

$\pi = 3{,}14$ sau $\pi = 3{,}141592$

Scanned with OKEN Scanner

Observaţie:

- Într-un calcul manual: $6\cdot\left(\frac{1}{3}\right)=\frac{6\cdot 1}{3}=\frac{6}{3}=2$

- La calculator: $6\cdot\left(\frac{1}{3}\right)=6\cdot 0{,}33=1{,}98$

sau $6\cdot\left(\frac{1}{3}\right)=6\cdot 0{,}333333=1{,}999998$

(funcţie de calculator)

Eroarea de calcul

Este de două feluri:

- de trunchiere şi
- de rotunjire

Eroarea de trunchiere apare din cauza trunchierii unui calcul numeric infinit.

Exemplu: la serii infinite convergente $S=\sum_{k=1}^{\infty} f(k)$

Dacă aproximăm pe S prin suma parţială $S_n=\sum_{k=1}^{n} f(k)$ (denumită seria trunchiată) atunci partea înlăturată (notată R_n) se numeşte restul seriei.

Avem: $\tilde{S}=S_n;\quad S=\tilde{S}+R_n$

$$e_{Sn}=S-\tilde{S}=R_n \qquad (5)$$

adică, eroarea de trunchiere obţinută conţinută în aproximaţia $\tilde{S}$ este restul seriei, R_n.

Eroarea de rotunjire este eroarea care apare datorită numărului limitat de cifre semnificative *pe care calculatorul le poate păstra* în *urma operaţiilor aritmetice* **în virgulă mobilă normalizată.**

Definiţie. Un număr este în virgulă mobilă dacă are forma $x=M\cdot b^E$, unde — M este mantisa (pozitivă sau negativă)

— b= baza sistemului de numeraţie (b=2, b=10,...)

Scanned with OKEN Scanner

— E este exponentul (pozitiv sau negativ)

Exemplu: dacă b=10, următoarele forme ale numărului x = 245,3 sunt în virgulă mobilă

$x=24,53 \cdot 10 = 2,453 \cdot 10^2 = 0,2453 \cdot 10^3 = 0,02453 \cdot 10^4 = 0,002453 \cdot 10^5 = 24530,0 \cdot 10^{-2}$

Definiție. Un număr x este în *virgulă mobilă, normalizat,* dacă este în virgulă mobilă, iar mantisa M verifică condițiile

$$\frac{1}{b} \le M < 1 \qquad (6)$$

Exemplu: numărul 245,3 în virgulă mobilă va avea forma normalizată dacă se scrie sub forma: $0,2453 \cdot 10^3$

Operații aritmetice cu numere în virgulă mobilă normalizate

Fie $x = M_x \cdot 10^{Ex}$, $y = M_y \cdot 10^{Ey}$

La înmulțire

$$Z = x \cdot y = M_x \cdot 10^{Ex} \cdot M_y 10^{Ey} = M_x \cdot M_y \cdot 10^{Ex+Ey}$$

- dacă $0,1 \le M_x M_y < 1$ atunci $\left.\begin{aligned} M_z &= M_x \cdot M_y \\ E_z &= E_x + E_y \end{aligned}\right\}$ (7)

- dacă $|M_x M_y| \ge 1$ atunci $\left.\begin{aligned} M_z &= \frac{M_x \cdot M_y}{10} \\ E_z &= E_x + E_y + 1 \end{aligned}\right\}$ (8)

- dacă $|M_x M_y| < 0,1$ atunci

$$M_z = M_x \cdot M_y \cdot 10, \quad E_z = E_x + E_y - 1$$

Exemplu:

$$x = 0,3184 \cdot 10^2; \quad y = 0,2512 \cdot 10^2$$

$$xy = z = 0,07998208 \cdot 10^4 = 0,7998208 \cdot 10^3$$

La împărțire

$$Z = \frac{x}{y} = \frac{M_x \cdot 10^{Ex}}{M_y \cdot 10^{Ey}} = \frac{M_x}{M_y} \cdot 10^{Ex-Ey}$$

Scanned with OKEN Scanner

- dacă $0,1 \le \left|\frac{M_x}{M_y}\right| < 1$ atunci $\frac{M_x}{M_y} = M_z ; E_z = E_x - E_y$ (9)

- dacă $\left|\frac{M_x}{M_y}\right| \ge 1$, atunci $\left.\begin{aligned} M_z &= \frac{M_x / M_y}{10} \\ E_z &= E_x - E_y + 1 \end{aligned}\right\}$ (10)

Exemplu:

$$x = 0,4251 \cdot 10^3 \quad \left| \quad \frac{x}{y} = 1,72454361 \cdot 10^1 = 0,172454361 \cdot 10^2 \right.$$
$$y = 0,2465,10^2$$

La adunare şi scădere

Presupunem că $E_x \ge E_y$. Avem

$$Z = x \pm y = M_x \cdot 10^{Ex} \pm M_y \cdot 10^{Ey} = \left(M_x \pm M_y \cdot 10^{Ey-Ex}\right) \cdot E^x$$

Notăm $\hat{M}_y = M_y \cdot 10^{E_y - E_x}$, avem: $z = \left(M_x \pm \hat{M}_y\right) 10^{Ex}$

- dacă $0,1 \le \left(M_x \pm \hat{M}_y\right) < 1$ atunci

$M_z = M_x \pm \hat{M}_y$, $E_z = E_x$

dacă $\left|M_x \pm \hat{M}_y\right| < 0,1$ atunci

$M_z = 10^n \cdot \left(M_x \pm \hat{M}_y\right)$, $E_z = E_{x-n}$

unde n este astfel încât: $0,1 \le |M_z| < 1$

Exemplu:

$x = 0,2844 \cdot 10^3$

$y = 0,4162 \cdot 10^1 = 0,004162 \cdot 10^3$

$x + y = z = 0,288562 \cdot 10^3$

Scanned with OKEN Scanner

Rotunjirea în operațiile aritmetice cu numere în virgulă mobilă normalizate

Fie $x = M_x \cdot 10^{Ex}$, $y = M_y \cdot 10^{Ey}$, unde M_x, M_y conțin d cifre.

În evaluarea $x \cdot y$, $\frac{x}{y}$, mantisa $M_z\left(z = x \cdot y, \quad z = \frac{x}{y}\right)$ poate conține $2 \cdot$ d cifre. Dacă în calculator nu se pot memora decât numere normalizate, având mantisa cu maximum d cifre, atunci este necesară rotunjirea mantisei M_z înainte de a memora pe z.

Fie $M_z = 0 \cdot \alpha_1 \alpha_2 \ldots \alpha_d \alpha_{d+1} \alpha_{d+2} \ldots \alpha_{2d}$

unde $\alpha_1 \alpha, \ldots \alpha_d$ *sunt cifrele cele mai semnificative* iar $\alpha_{d+1} \alpha_{d+2} \ldots \alpha_{2d}$ sunt cifrele cele mai puțin semnificative. Atunci

$$Z = M_z \cdot 10^{Ez} = \left(0 \cdot \alpha_1 \alpha_2 \ldots \alpha_d + 0 \cdot 000 \ldots 0 \alpha_{d+1} \alpha_{d+2} \ldots \alpha_{2d}\right) \cdot 10^{Ez} =$$
$$= \left(0 \cdot \alpha_1 \alpha_2 \ldots \alpha_d + 0 \cdot \alpha_{d+1} \alpha_{d+2} \ldots \alpha_{2d} \cdot 10^{-d}\right) \cdot 10^{Ez} =$$
$$= M \cdot 10^{Ez} + m \cdot 10^{Ez-d}$$

unde $M = 0 \cdot \alpha_1 \alpha_2 \ldots \alpha_d$, iar $m = 0 \cdot \alpha_{d+1} \alpha_{d+2} \ldots \alpha_{2d}$

Regula de rotunjire

- Dacă $|m| < 0 \cdot 5$, atunci $|\tilde{Z}| = |M| \cdot 10^{Ez}$, denumită rotunjire prin lipsă.
- Dacă $|m| \geq 0 \cdot 5$, atunci

  $|\tilde{Z}| = |M| \cdot 10^{Ez} + 10^{Ez-d} = \left(|M| + 10^{-d}\right) \cdot 10^{Ez}$, aceasta fiind rotunjirea prin adaos.

Exemplu: Admitem d = 4 iar

$Z = -0.21567891 \cdot 10^{+5} = [-0,2156 + (-0.0000 7891)] \cdot 10^5 =$
$= -0,2156 \cdot 10^5 + \left(-0.7891 \cdot 10^{+1}\right)$, avem: $M = -0.2156$ și $m = -0.7891$. Deoarece $|m| = 0.7891 > 0.5$ rezultă că

$\tilde{Z} = -\left(0.2156 + 10^{-4}\right) \cdot 10^5 = -\left(0.2156 + 0.0001\right) \cdot 10^5 = -0.2157 \cdot 10^5$

(rotunjire prin adaus)

Scanned with OKEN Scanner

Concluzii

Scriere

- $Z = \underbrace{M_z}_{\text{cu 2d cifre semnificative}} \cdot 10^{Ez} = M \cdot 10^{Ez} + m \cdot 10^{E-d}$ (11)

unde $M = 0.\ \alpha_1\alpha_2 \ldots \alpha_d$, $m = 0.\alpha_{d+1}\alpha_{d+2} \ldots \alpha_{2d}$

Rotunjire

- $|m| < 0.5 \Rightarrow |\tilde{Z}| = |M| \cdot 10^{Ez}$ (rotunjire prin lipsă) (12)
- $|m| > 0.5 \Rightarrow |\tilde{Z}| = (|M| + 10^{-d}) \cdot 10^{Ez}$ (13)

(rotunjire prin adaus)

Margine superioară pentru eroarea absolută de rotunjire

Fie Z un număr în virgulă mobilă normalizat

$\tilde{Z}$ – valoarea obținută prin rotunjire;

$e_{\tilde{z}} = z - \tilde{z}$, $|e_{\tilde{z}}| = |z - \tilde{z}|$ = eroarea absolută de rotunjire

$Z = M \cdot 10^{Ez} + m \cdot 10^{Ez-d}$

- dacă $|m| < 0.5$ atunci $\tilde{Z} = M \cdot 10^{Ez}$

$$|e_{\tilde{z}}| = |M \cdot 10^{Ez} + m \cdot 10^{Ez-d} - M \cdot 10^{Ez}| = |m| \cdot 10^{Ez-d} < 0.5 \cdot 10^{Ez-d}$$

Concluzie $|e_{\tilde{z}}| \le \underbrace{0.5 \cdot 10^{Ez-d}}_{\text{margine superioară pentru eroarea absolută } (|e_{\tilde{z}}|) \text{ de rotunjire}}$

- dacă $|m| \ge 0.5$ şi $z > 0 \underset{(13)}{\Rightarrow} \tilde{Z} = (M + 10^{-d}) \cdot 10^{Ez}$ şi

$$|e_{\tilde{z}}| = |z - \tilde{z}| = |M \cdot 10^{Ez} + m \cdot 10^{Ez-d} - M \cdot 10^{Ez} - 10^{Ez-d}| =$$

$$= |m - 1| \cdot 10^{Ez-d} \le 0.5 \cdot 10^{Ez-d}$$

- dacă $|m| \ge 0.5$ şi $z < 0$ (M şi m sunt negative)

notăm: $-M = M'$, $-m = m'$

$$|e_{\tilde{z}}| = |-M' \cdot 10^{Ez} - m' \cdot 10^{Ez-d} + (M' \cdot 10^{Ez-d})| =$$

Scanned with OKEN Scanner

$$= |1 - m'| \cdot 10^{Ez-d} = |1 + m| \cdot 10^{Ez-d} \leq 0.5 \cdot 10^{Ez-d}$$

Concluzie. Pentru orice z avem

$$|e_{\tilde{z}}| \leq \underbrace{0.5 \cdot 10^{Ez-d}}_{\text{reprezintă marginea superioară pentru eroarea absolută de rotunjire}}$$

Margine superioară pentru eroarea relativă de rotunjire în valoare absolută

Fie Z un număr în virgulă mobilă

$\tilde{Z}$ – valoarea obținută prin rotunjire;

$e_{\tilde{z}} = z - \tilde{z}$, eroarea conținută în $\tilde{z}$

$r_{\tilde{z}} = \frac{e_{\tilde{z}}}{\tilde{z}}$ eroarea relativă conținută în $\tilde{z}$

Avem:

$$|r_{\tilde{z}}| = \left|\frac{e_{\tilde{z}}}{\tilde{z}}\right| = \frac{|e_{\tilde{z}}|}{|\tilde{z}|} \leq \frac{0.510^{Ez-d}}{|\tilde{z}|}$$

dar $|\tilde{z}| \geq |M| \cdot 10^{Ez} \geq 0.1 \cdot 10^{Ez}$ Rezultă

$$|r_{\tilde{z}}| \leq \frac{0.5.10^{Ez-d}}{0.1.10^{Ez}} = 5 \cdot 10^{-d}$$

În concluzie $|r_{\tilde{z}}| \leq 5 \cdot 10^{-d}$ (14)

Propagarea erorilor prin operațiile aritmetice

1. **Fără eroare de rotunjire**

Fie x,y; $\tilde{x}, \tilde{y}$

$e_{\tilde{x}} = x - \tilde{x}$; $e_{\tilde{y}} = y - \tilde{y}$

- **la adunare**

$$e_{\tilde{x}+\tilde{y}} = (x + y) - (\tilde{x} + \tilde{y}) = (x - \tilde{x}) + (y - \tilde{y}) = e_{\tilde{x}} + e_{\tilde{y}}$$

$$\Rightarrow e_{\tilde{x}+\tilde{y}} = e_{\tilde{x}} + e_{\tilde{y}} \quad (15)$$

Scanned with OKEN Scanner

Scăderea

$$e_{\tilde{x}-\tilde{y}} = (x-y)-(\tilde{x}-\tilde{y}) = (x-\tilde{x})-(y-\tilde{y}) = e_{\tilde{x}} - e_{\tilde{y}} \Rightarrow e_{\tilde{\tilde{x}}-\tilde{y}} = e_{\tilde{x}} - e_{\tilde{y}} \quad (16)$$

Înmulțirea

$$e_{\tilde{x}\cdot\tilde{z}} = (x\cdot y)-(\tilde{x}\cdot\tilde{y}) = (\tilde{x}+e_{\tilde{x}})\cdot(\tilde{y}+e_{\tilde{y}})-\tilde{x}\,\tilde{y} =$$

$$= \tilde{x}\,\tilde{y} + e_{\tilde{x}}\cdot\tilde{y} + \tilde{x}\cdot e_{\tilde{y}} \,\big|\cdot e_{\tilde{x}}\cdot e_{\tilde{y}} - \tilde{x}\cdot\tilde{y}$$

$$\Rightarrow \quad e_{\tilde{x}\cdot\tilde{y}} = e_{\tilde{x}}\cdot\tilde{y} + \tilde{x}\cdot e_{\tilde{y}} \qquad (18)$$

unde $e_{\tilde{x}}\cdot e_{\tilde{y}} = 0$

Împărțirea

$$e_{\tilde{x}/\tilde{y}} = \frac{x}{y} - \frac{\tilde{x}}{\tilde{y}} = \frac{x\cdot\tilde{y}-\tilde{x}\cdot y}{y\,\tilde{y}} = \frac{(\tilde{x}+e_{\tilde{x}})\,\tilde{y}-\tilde{x}(\tilde{y}+e_{\tilde{y}})}{(\tilde{y}+e_{\tilde{y}})\cdot\tilde{y}}$$

$$l_{\tilde{x}/\tilde{y}} = \frac{(\tilde{x}\;\tilde{y})+e_{\tilde{x}}\cdot y-\tilde{x}\,\tilde{y}-\tilde{x}\,e_{\tilde{y}}}{\tilde{y}^2\left(1+\frac{e_{\tilde{y}}}{\tilde{y}}\right)} \quad \Rightarrow \quad e_{\tilde{x}/\tilde{y}} = \frac{e_{\tilde{x}}\cdot\tilde{y}-\tilde{x}\cdot e_{\tilde{y}}}{\tilde{y}^2}$$

unde $e_{\tilde{y}}/\tilde{y} = 0$

2. Propagarea erorilor prin operațiile aritmetice, cu considerarea erorii de rotunjire

Adunarea

$$e_{\tilde{\tilde{x}}+\tilde{y}} = e_{\tilde{x}} + e_{\tilde{y}} + \overline{\alpha} \qquad (16)$$

unde $\overline{\alpha}$ este eroarea de rotunjire conținută în $\tilde{x}+\tilde{y}$.

Scăderea

$$e_{\tilde{\tilde{x}}-\tilde{y}} = e_{\tilde{x}} - e_{\tilde{y}} + \overline{\delta} \qquad (17)$$

unde $\overline{\delta}$ este eroarea de rotunjire conținută în $\tilde{x}-\tilde{y}$

Scanned with OKEN Scanner

Înmulțirea

$$e_{\tilde{x}\cdot\tilde{y}} = e_{\tilde{x}}\cdot\tilde{y} + \tilde{x}\cdot e_{\tilde{y}} + \overline{\mu} \tag{18}$$

unde $\overline{\mu}$ este eroarea de rotunjire conținută în $\tilde{x}\cdot\tilde{y}$

Împărțirea

$$e_{\tilde{x}/\tilde{y}} = \frac{e_{\tilde{x}}\cdot\tilde{y} - \tilde{x}\cdot e_{\tilde{y}}}{\tilde{y}^2} + \overline{\delta} \tag{19}$$

unde $\overline{\delta}$ este eroarea de rotunjire conținută în $\tilde{x}/\tilde{y}$.

Observație
Eroarea de rotunjire este independentă de operația aritmetică

3. Propagarea erorilor relative prin operațiile aritmetice

Fie $x, y; \tilde{x}, \tilde{y}; e_{\tilde{x}}, e_{\tilde{y}}; r_{\tilde{x}} = \frac{e_{\tilde{x}}}{\tilde{x}}, r_{\tilde{y}} = \frac{e_{\tilde{y}}}{\tilde{y}}$

Adunarea

$$r_{\tilde{x}+\tilde{y}} = \frac{e_{\tilde{x}} + \tilde{y}}{\tilde{x}+\tilde{y}} = \frac{e_{\tilde{x}} + e_{\tilde{y}} + \overline{\alpha}}{\tilde{x}+\tilde{y}} = \frac{r_{\tilde{x}}\cdot\tilde{x} + r_{\tilde{y}}\,\tilde{y} + \overline{\alpha}}{\tilde{x}+\tilde{y}} = r_{\tilde{x}}\cdot\frac{\tilde{x}}{\tilde{x}+\tilde{y}} + r_{\tilde{y}}\cdot\frac{\tilde{y}}{\tilde{x}+\tilde{y}} +$$

$$+ \underbrace{\frac{\overline{\alpha}}{\tilde{x}+\tilde{y}}}_{\alpha''} \Rightarrow r_{\tilde{x}+\tilde{y}} = r_{\tilde{x}}\cdot\frac{\tilde{x}}{\tilde{x}+\tilde{y}} + r_{\tilde{y}}\cdot\frac{\tilde{y}}{\tilde{x}+\tilde{y}} + \alpha \tag{20}$$

Scăderea

$$r_{\tilde{x}-\tilde{y}} = \frac{e_{\tilde{x}} - \tilde{y}}{\tilde{x}-\tilde{y}} = \frac{e_{\tilde{x}} - e_{\tilde{y}} + \overline{\delta}}{\tilde{x}-\tilde{y}} = \frac{r_{\tilde{x}}\cdot\tilde{x} - r_{\tilde{y}}\,\tilde{y} + \overline{\delta}}{\tilde{x}-\tilde{y}} = r_{\tilde{x}}\cdot\frac{\tilde{x}}{\tilde{x}-\tilde{y}} - r_{\tilde{y}}\cdot\frac{\tilde{y}}{\tilde{x}-\tilde{y}} +$$

Scanned with OKEN Scanner

$$+\underbrace{\frac{\overline{\delta}}{\tilde{x}-\tilde{y}}}_{\sigma} \Rightarrow r_{\tilde{x}-\tilde{y}} = r_{\tilde{x}} \cdot \frac{\tilde{x}}{\tilde{x}-\tilde{y}} - r_{\tilde{y}} \cdot \frac{\tilde{y}}{\tilde{x}-\tilde{y}} + \delta \qquad (21)$$

Înmulțirea

$$r_{\tilde{x}\cdot\tilde{y}} \underline{\underline{\text{def}}} \frac{l_{\tilde{x}} \cdot \tilde{y}}{\tilde{x}\cdot\tilde{y}} = \frac{e_{\tilde{x}} \cdot \tilde{y} + \tilde{x} \cdot e_{\tilde{y}} + \overline{\mu}}{\tilde{x}\cdot\tilde{y}} = r_{\tilde{x}} + r_{\tilde{y}} + \underbrace{\frac{\overline{\mu}}{\tilde{x}\cdot\tilde{y}}}_{\mu}$$

$$\Rightarrow \quad r_{\tilde{x}\cdot\tilde{y}} = r_{\tilde{x}} + r_{\tilde{y}} + \mu \qquad (22)$$

Împărțirea

$$r_{\tilde{x}/\tilde{y}} = \frac{e_{\tilde{x}}/\tilde{y}}{\tilde{x}/\tilde{y}} = \frac{\dfrac{e_{\tilde{x}} \cdot \tilde{y} - \tilde{x} \cdot e_{\tilde{y}} + \overline{\delta}}{\tilde{y}^2}}{\tilde{x}/\tilde{y}} = \frac{e_{\tilde{x}}}{\tilde{x}} - \frac{e_{\tilde{y}}}{\tilde{y}} + \frac{\overline{\delta}}{\tilde{x}/\tilde{y}}$$

$$\Rightarrow \quad r_{\tilde{x}/\tilde{y}} = r_{\tilde{x}} - r_{\tilde{y}} + \delta \qquad (23)$$

Observație : $\alpha, \sigma, \mu, \delta$ sunt **erorile relative** de rotunjire conținute respectiv în $(\tilde{x}+\tilde{y})$; $(\tilde{x}-\tilde{y})$; $\tilde{x}\cdot\tilde{y}$; $\tilde{x}/\tilde{y}$.

Observație : Efectul propagării erorilor este foarte important și trebuie să fie supravegheat, deoarece uneori poate fi dezastruos.

Analiza propagării erorii într-o expresie aritmetică

Să se calculeze suma : $\tilde{S} = \tilde{a}_0 + \tilde{a}_1 + \tilde{a}_2 + \tilde{a}_3$

Suma se va calcula astfel : $\tilde{S} = [(\tilde{a}_0 + \tilde{a}_1) + \tilde{a}_2] + \tilde{a}_3$

Notăm : $\tilde{S}_1 = \tilde{a}_0 + \tilde{a}_1$, $\tilde{S}_2 = \tilde{S}_1 + \tilde{a}_2$, $\tilde{S}_3 = \tilde{S}_2 + \tilde{a}_3 = \tilde{S}$

Notăm cu :

- r_0, r_1, r_2, r_3 erorile relative conținute în $\tilde{a}_0$, $\tilde{a}_1$, $\tilde{a}_2$ și respectiv $\tilde{a}_3$

  $\alpha_1, \alpha_2, \alpha_3$ erorile relative de rotunjire conținute în $\tilde{S}_1$, $\tilde{S}_2$, $\tilde{S}_3$.

Scanned with OKEN Scanner

Din (20) $\Big|\ r_{\tilde{x}+\tilde{y}} = r_{\tilde{x}} \cdot \frac{\tilde{x}}{\tilde{x}+\tilde{y}} + r_{\tilde{y}} \cdot \frac{\tilde{y}}{\tilde{x}+\tilde{y}} + \alpha)$

$\tilde{S}_1 = \tilde{a}_0 + \tilde{a}_1$

$r_{S_1} = r_0 \cdot \frac{\tilde{a}_0}{\tilde{a}_0+\tilde{a}_1} + r_1 \cdot \frac{\tilde{a}_1}{\tilde{a}_0+\tilde{a}_1} + \alpha_1 = r_0 \cdot \frac{\tilde{a}_0}{\tilde{S}_1} + r_1 \cdot \frac{\tilde{a}_1}{\tilde{S}_1} + \alpha_1$

$\tilde{S}_2 = \tilde{S}_1 + \tilde{a}_2$

$r_{S_2} = r_{\tilde{S}_1} \cdot \frac{\tilde{S}_1}{\tilde{S}_1+\tilde{a}_2} + r_2 \cdot \frac{\tilde{a}_2}{\tilde{S}_1+\tilde{a}_2} + \alpha_2 \Rightarrow$

$r_{S_2} = r_{\tilde{S}_1} \cdot \frac{\tilde{S}_1}{\tilde{S}_2} + r_2 \cdot \frac{\tilde{a}_2}{\tilde{S}_2} + \alpha_2 = \left[\left(r_0 \frac{\tilde{a}_0}{\tilde{S}_1} + r_1 \frac{a_1}{\tilde{S}_1} + \alpha_1 \right) \cdot \frac{\tilde{S}_1}{\tilde{S}_2} + r_2 \frac{\tilde{a}_2}{\tilde{S}_2} + \alpha_2 \right] =$

$= \frac{r_0 \tilde{a}_0}{\tilde{S}_2} + \frac{r_1 \tilde{a}_1}{\tilde{S}_2} + \alpha_1 \frac{\tilde{S}_1}{\tilde{S}_2} + \frac{r_2 \tilde{a}_2}{\tilde{S}_2} + \alpha_2$

$\tilde{S} = \tilde{S}_3 = \tilde{S}_2 + \tilde{a}_3$;

$\tilde{r}_S = r_{\tilde{S}_2} \cdot \frac{\tilde{S}_2}{\tilde{S}} + r_3 \cdot \frac{\tilde{a}_3}{\tilde{S}} + \alpha_3 = \frac{r_0 \tilde{a}_0}{\tilde{S}} + \frac{r_1 \tilde{a}_1}{\tilde{S}} + \frac{r_2 \tilde{a}_2}{\tilde{S}} + \frac{r_3 \tilde{a}_3}{\tilde{S}} +$

$+ \alpha_1 \frac{\tilde{S}_1}{\tilde{S}} + \frac{\alpha \tilde{S}_2}{\tilde{S}} + \alpha_3 \frac{\tilde{S}_3}{S}$

$r_{\tilde{S}} = \frac{\sum_{i=0}^{3} r_i \tilde{a}_i + \sum_{i=1}^{3} \alpha_i \tilde{S}_i}{\tilde{S}}$

Dar :

$$r_{\tilde{S}} = \frac{e_{\tilde{S}}}{\tilde{S}} \quad \Rightarrow \quad e_{\tilde{S}} = r_{\tilde{S}}\ \tilde{S} \text{ sau}$$

$$e_{\tilde{S}} = \sum_{i=0}^{3} r_i\ \tilde{a}_i + \sum_{i=1}^{3} \alpha_i\ \tilde{S}_i \quad (24)$$

unde : $|\alpha_i| \le 5 \cdot 10^{-d}$, $i = 1, 2, 3.$

Dacă există $K \in \Re\ a.\hat{i}. |r_i| \le K \cdot 10^{-d}$, $i = \overline{0, 3}$ atunci

Scanned with OKEN Scanner

$$\left|e_{\tilde{S}}\right| \le \left(K \cdot \sum_{i=0}^{3} \left|\tilde{a}_i\right| + 5 \sum_{i=1}^{3} \left|\tilde{S}_i\right| \right) \cdot 10^{-d} \qquad (25)$$

Observație : Dacă numerele $\tilde{a}_i$, $i = \overline{0,3}$ au toate același semn, atunci sumele parțiale $\tilde{S}_1$, $\tilde{S}_2$, $\tilde{S}_3$ sunt mai mici în cazul

$|\tilde{a}_0| \le |\tilde{a}_1| \le |\tilde{a}_2| \le |\tilde{a}_3|$ (ordine crescătoare), decât în cazul

$|\tilde{a}_0| \ge |\tilde{a}_1| \ge |\tilde{a}_2| \ge |\tilde{a}_3|$ (ordine descrescătoare). Rezultă că marginea erorii este mai mică în primul caz.

Exemplu :

$\tilde{a}_0 = 1$, $\tilde{a}_1 = 2$, $\tilde{a}_2 = 3$, $\tilde{a}_3 = 4$

$1 \le 2 \le 3 \le 4$

$\tilde{S}_1 = 1 + 2 + 3$; $\tilde{S}_2 = 3 + 3 = 6$; $\tilde{S}_3 = 6 + 4 = 10$

$$\sum_{i=1}^{3} \tilde{S}_i = 19$$

Invers : 4, 3, 2, 1

$S_1 = 7$, $S_2 = 9$, $S_3 = 10$

$$\sum_{i=0}^{3} S_i = 7 + 9 + 10 = 26$$

Concluzie :

Generalizând, dacă avem de adunat un şir de numere, toate de acelaşi semn, **este bine** să adunăm în ordinea crescătoare a valorii lor absolute.

Eroarea de metodă

Este eroarea introdusă **de metoda numerică folosită** şi poate fi micşorată prin alegerea convenabilă a acesteia.

Aplicație. Să se calculeze valoarea unui polinom într-un punct. Fie polinomul

$P_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n$ şi fie P valoarea sa în punctul $\tilde{x}$.

Scanned with OKEN Scanner

Metoda întâia :

$\mathbf{P} = \left(\left(\left(a_0\tilde{x}^n + a_1\tilde{x}^{n-1}\right) + a_2\tilde{x}^{n-2}\right) + \ldots + a_{n-i}\tilde{x}\right) + a_n$, unde calculul lui $a_0\tilde{x}^n$ necesită n înmulțiri, $a_1\tilde{x}^{n-1}$ necesită (n–1) înmulțiri etc. Sunt necesare: (1+2+…+n) înmulțiri $\left(1+2+\ldots+n = \frac{n(n+1)}{2}\right)$;

- **Rezultă** $\frac{n(n+1)}{2}$ înmulțiri şi n adunări pentru a calcula P.

Metoda a doua (Schema lui Horner)

$\mathbf{P} = \left(\ldots\left(\left(\left(a_0\right)\tilde{x} + a_1\right)\tilde{x} + a_2\right)\tilde{x} + \ldots\right)\tilde{x} + a_n$,

Avem:

$P = a_0$ şi $P = P\cdot\ \tilde{x} + a_i$, $i = \overline{1,n}$, adică sunt necesare n înmulțiri şi n adunări pentru a calcula P.

Concluzie Deoarece $\frac{n(n+1)}{2} \geq n$, pentru $n \geq 1$ rezultă că metoda a doua este mai bună decât prima, deoarece necesită mai puține înmulțiri, iar eroarea introdusă de ea va fi mai mică.

Clasificarea metodelor de calcul numeric din punctul de vedere al exactității soluției

Avem : **metode exacte**, care teoretic furnizează o soluție exactă
metode aproximative (interative sau neinterative) care teoretic furnizează o soluție aproximativă.

Observație. Practic la calculator şi metodele exacte pot produce tot soluții aproximative, din cauza rotunjirilor, a trunchierilor şi a propagării erorilor.

Probleme propuse

1. Să se analizeze propagarea erorii în expresia
$P = \tilde{a}_0\cdot\tilde{a}_1\cdot\tilde{a}_2\cdot\tilde{a}_3$ cu $\tilde{r}_0, \tilde{r}_1, \tilde{r}_2, \tilde{r}_3$ erorile relative conținute în $\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3$

2. Să se analizeze propagarea erorii în $P_2(x) =$
$= \tilde{a}_0\cdot\tilde{x}^2 + \tilde{a}_1\cdot\tilde{x} + \tilde{a}_2$, unde $\tilde{r}_0, \tilde{r}_1, \tilde{r}_2, \tilde{r}_3, \tilde{r}_x$ sunt erorile relative în $\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \tilde{a}_3$ şi $\tilde{x}$.

Scanned with OKEN Scanner

ANEXA 4

PROBABILITĂȚI ŞI REPARTIȚII TABELATE

Tabela I — Valorile probabilității $P(X \geq K)$ ale unui eveniment rar, care urmează legea Poisson

Tabela II — Densitatea de repartiție $n(x; 0,1)$ a legii normale normate

Tabela III — Funcția normată a lui Laplace $\phi(t)$

Tabela IV — Funcția normată a lui Laplace dublată $2 \cdot \phi(t)$

Tabela V — Valorile x_ε^2 în funcție de probabilitatea $P\left(x^2 > x_\varepsilon^2\right)$, de numărul k al gradelor de libertate ale repartiției x^2

Tabela VI — Probabilitatea $F(t;n)$ pentru repartiția Student, care depinde de t şi de numărul n al gradelor de libertate

Tabela VII — Limitele intervalului de încredere 95% pentru probabilitatea p din distribuția binominală, estimată cu frecvența relativă $\frac{x}{n}$

Tabela VIII — Valorile de 5% şi 1% ale limitelor abaterii variabilei F în funcție de gradele de libertate m şi n.

Tabela IX — Probabilitățile $P_n(\lambda) = K(\lambda)$ ale repartiției lui Kolmogarov.

Scanned with OKEN Scanner

Tabela I

Valorile probabilității $P(X \geq k) = \sum_{x=k}^{\infty} e^{-a} \frac{a^x}{x!}$ ale unui eveniment rar, are urmează legea Poisson[1]

k	a=0,1	a=1	a=5	a=10	a=15	a=20
0	1,000000	1,000000	1,000000	1,000000	—	—
5	—	0,003660	0,559510	0,970750	0.999140	0.991930
10	—	—	0,031328	0,542070	0,930150	0,995000
15	—	—	0,000226	0,083458	0,534350	0,895140
20	—	—	—	0,034543	0,124780	0,529740
25	—	—	—	0,000047	0,111650	0,156770
30	—	—	—	—	0,000418	0,021818
35	—	—	—	—	0,000007	0,014890
40	—	—	—	—	—	0,000532
45	—	—	—	—	—	0,000001

[1] X reprezintă numărul aparițiilor evenimentului, iar *a* numărul mediu al acestor apariții, deoarece M(X) = *a* pentru distribuția lui Poisson.

Scanned with OKEN Scanner

Densitatea de repartiție $n(x;\,0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ **a legii normale normate**

x	0	1	2	3	4	5	6	7	8	9
0,0	3989^{-4}	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970^{-4}	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910^{-4}	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814^{-4}	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683^{-4}	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521^{-4}	3503	3485	3476	3448	3429	3410	3391	3372	3352
0,6	3332^{-4}	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123^{-4}	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897^{-4}	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661^{-4}	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420^{-4}	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179^{-4}	2155	2131	2107	2083	2059	2036	2012	1978	1965
1,2	1942^{-4}	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714^{-4}	1691	1669	1647	1426	1604	1582	1561	1539	1518
1,4	1497^{-4}	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295^{-4}	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109^{-4}	1092	1074	1057	1040	1023	1006	9893^{-5}	9728	9566
1,7	9405^{-5}	9246	9089	8933	8780	8628	8478	8329	8183	8038
1,8	7985^{-5}	7754	7614	7477	7341	7206	7074	6943	6814	6687
1,9	6562^{-5}	6438	6316	6195	6077	5959	5844	5730	5618	5508
2,0	5399^{-5}	5292	5186	5082	4980	4879	4780	4682	4586	4491
2,1	4398^{-5}	4307	4217	4128	4041	3955	3871	3788	3706	3626
2,2	3547^{-5}	3470	3394	3319	3246	3174	3103	3034	2965	2898
2,3	2883^{-5}	2768	2705	2643	2582	2522	2463	2406	2349	2294
2,4	2239^{-5}	2186	2134	2083	2033	1984	1936	1888	1842	1797
2,5	1753^{-5}	1709	1667	1625	1585	1545	1506	1468	1431	1394
2,6	1358^{-5}	1323	1289	1256	1223	1191	1160	1130	1100	1071
2,7	1042^{-5}	1014	9871^{-6}	9606	9347	9094	8846	8605	8370	8140
2,8	7915^{-6}	7697	7483	7274	7071	6873	6679	6491	6307	6127
2,9	5953^{-6}	5782	5616	5454	5296	5143	4993	4847	4705	4567
3,0	4432^{-6}	4301	4173	4049	3928	3810	3695	3584	3475	3370
3,1	3267^{-6}	3167	3070	2975	2884	2794	2707	2623	2541	2461
3,2	2384^{-6}	2309	2236	2165	2096	2029	1964	1901	1840	1780
3,3	1723^{-6}	1667	1612	1560	1508	1459	1411	1364	1319	1275
3,4	1232^{-6}	1191	1151	1112	1075	1038	1003	9689^{-7}	9358	9037
3,5	8727^{-7}	8426	8135	7853	7581	7317	7061	6814	6575	6343
3,6	6119^{-7}	5902	5693	5490	5294	5105	4921	4744	4573	4408
3,7	4248^{-7}	4093	3944	3800	3661	3526	3396	3271	3149	3032
3,8	2919^{-7}	2810	2705	2604	2506	2411	2320	2232	2147	2065
3,9	1987^{-7}	1940	1887	1766	1698	1633	1569	1508	1449	1393
4,0	1338^{-7}	1286	1235	1186	1140	1094	1051	1009	9687^{-8}	9299
4,1	8926^{-8}	8567	8222	7890	7570	7263	6967	6683	6410	6147
4,2	5894^{-8}	5652	5418	5194	4979	4772	4573	4382	4199	4023
4,3	3854^{-8}	3691	3535	3386	3242	3104	2972	2845	2723	2606
4,4	2494^{-8}	2387	2284	2185	2090	1999	1912	1829	1749	1672
4,5	1598^{-8}	1528	1461	1396	1334	1275	1218	1164	1112	1062
4,6	1014^{-8}	9684^{-9}	9248	8830	8430	8047	7681	7331	6996	6676
4,7	6370^{-9}	6077	5797	5530	5274	5030	4796	4573	4360	4156
4,8	3961^{-9}	3775	3598	3428	3267	3112	2965	2824	2690	2561
4,9	2439^{-9}	2322	2211	2105	2003	1907	1814	1767	1643	1526

Notă. Exponentul negativ se referă la factorul 10, deci $1987^{-7} = 1987 \times 10^{-7} = 0,0001987$. Dacă un număr nu are exponent, i se dă exponentul numărului celui mai apropiat din stînga. De pildă la 4,02 corespunde numărul 1235 care nu are exponent, de aceea îi dăm exponentul -7, adică valoarea ordonatei $n(4,02\,;\,1,0) = 1235^{-7} = 0,0001235$.

Scanned with OKEN Scanner

Funcția normală a lui Laplace $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}} dx.$

t	0	1	2	3	4	5	6	7	8	9
0,0	0,0 000	0,40	0,80	120	160	199	239	279*	319	350
0,1	398	438	478	517	557	596	636	675	714	753
0,2	793	832	871	910	948	987	026*	064*	103*	141*
0,3	0,1 179	217	255	293	331	368	406	443	480	517
0,4	554	591	628	664	700	736	772	808	844	879
0,5	915	950	985	019*	054*	088*	123*	157*	190*	224*
0,6	0,2 257	291	324	357	389	422	454	486	517	549
0,7	580	611	642	673	703	734	764	794	823	852
0,8	881	910	939	967	995	023*	051*	078*	106*	133*
0,9	0,3 159	186	212	238	264	289	315	340	365	389
1,0	413	437	461	485	508	531	554	577	599	621
1,1	643	665	686	708	729	749	770	790	810	830
1,2	849	869	888	907	925	944	962	980	997	015*
1,3	0,4 032	049	066	082	099	115	131	147	162	177
1,4	192	207	222	236	251	265	279	292	306	319
1,5	332	345	357	370	382	394	406	418	429	441
1,6	452	463	474	484	495	505	515	525	535	545
1,7	554	564	573	582	591	599	608	616	625	633
1,8	641	649	656	664	671	678	686	693	699	706
1,9	713	719	726	732	738	744	750	756	761	767
2,0	772	778	783	788	793	798	803	808	812	817
2,1	821	826	830	834	838	842	846	850	854	857
2,2	860	864	867	871	874	877	880	883	886	889
2,3	892	895	898	900	903	906	908	911	913	915
2,4	918	920	922	924	926	928	930	932	934	936
2,5	0,4 937	939	941	942	944	946	947	949	950	952
2,6	953	954	956	957	958	959	960	962	963	964
2,7	965	966	967	968	969	970	971	971	972	973
2,8	974	975	975	976	977	978	978	979	980	980
2,9	981	981	982	983	983	984	984	985	985	986
3,0	986	986	987	987	988	988	988	989	989	989
3,1	990	990	990	991	991	991	992	992	992	992
3,2	993	993	993	993	994	994	994	994	994	994
3,3	995	995	995	995	995	995	996	996	996	996
3,4	996	996	996	996	997	997	997	997	997	997
3,5	997	997	997	997	997	998	998	998	998	998
3,6	998	998	998	998	998	998	998	998	998	998
3,7	998	998	999	999	999	999	999	999	999	999
3,8	999	999	999	999	999	999	999	999	999	999
3,9	999	999	999	999	999	999	999	999	999	999
4,0	999	999	999	999	999	999	999	999	999	999
4,1	999	999	999	999	999	999	999	999	999	999
4,2	999	999	999	999	999	999	999	999	999	999
4,3	999	999	999	999	999	999	999	999	999	999
4,4	999	999	999	999	999	999	999	999	999	999
4,5	999	999	999	999	999	999	999	999	999	999
5,0	999									

Scanned with OKEN Scanner

t	0	1	2	3	4	5	6	7	8	9
0,0	0,0 0000	0798	1596	2393	3191	3988	4784	5581	6376	7171
0,1	7966	8759	9552	0343*	1134*	1924*	2712*	3499*	4285*	5069*
0,2	0,1 5852	6633	7413	8191	8967	9741	0514*	1284*	2052*	2818*
0,3	0,2 3582	4344	5103	5860	6614	7366	8115	8862	9605	0346*
0,4	0,3 1084	1819	2552	3280	4006	4729	5448	6164	6877	7587
0,5	8292	8995	9694	0389*	1080*	1768*	2452*	3132*	3809*	4481*
0,6	0,4 5149	5814	6474	7131	7783	8431	9075	9714	0350*	0981*
0,7	0,5 1607	2230	2848	3461	4070	4675	5275	5870	6161	7047
0,8	7629	8206	8778	9346	9909	0468*	1021*	1570*	2114*	2653*
0,9	0,6 3188	3718	4243	4763	5278	5789	6294	6795	7291	7783
1,0	8269	8750	9227	9699	0166*	0628*	1086*	1538*	1986*	2429*
1,1	0,7 2867	3300	3729	4152	4571	4986	5395	5800	6200	6595
1,2	6986	7372	7754	8130	8502	8870	9233	9592	9945	0295*
1,3	0,8 0640	0980	1316	1648	1975	2298	2617	2931	3241	3547
1,4	3849	4146	4439	4728	5013	5294	5571	5844	6113	6378
1,5	6639	6896	7149	7398	7644	7886	8124	8358	8589	8817
1,6	9040	9260	9477	9690	9899	0106*	0309*	0508*	0704*	0897*
1,7	0,9 1078	1273	1457	1637	1814	1988	2159	2327	2492	2655
1,8	2814	2970	3124	3275	3423	3569	3711	3852	3989	4124
1,9	4257	4387	4514	4639	4762	4882	5000	5116	5230	5341
2,0	5450	5557	5662	5764	5865	5964	6060	6155	6247	6338
2,1	6427	6514	6599	6682	6765	6844	6923	6999	7074	7148
2,2	7219	7289	7358	7452	7491	7555	7618	7679	7739	7798
2,3	7855	7911	7966	8019	8072	8123	8172	8221	8269	8315
2,4	8360	8405	8448	8490	8531	8571	8611	8649	8686	8723
2,5	8758	8793	8826	8859	8891	8923	8953	8983	9012	8040
2,6	9068	9095	9121	9146	9171	9195	9219	9241	9263	9285
2,7	9307	9327	9347	9367	9386	9404	9422	9439	9456	9473
2,8	9489	9505	9520	9535	9549	9563	9576	9590	9602	9615
2,9	9627	9639	9650	9661	9672	9682	9692	9702	9712	9721
3,0	9730	9739	9747	9755	9763	9771	9779	9786	9793	9800
3,1	9806	9813	9819	9825	9831	9837	9842	9848	9853	9858
3,2	9863	9867	9872	9876	9880	9885	9889	9892	9896	9900
3,3	9903	9907	9910	9913	9916	9919	9922	9925	9928	9930
3,4	9933	9936	9937	9940	9942	9944	9946	9948	9950	9952
3,5	9953	9955	9957	9958	9960	9961	9963	9964	9966	9967
3,6	9968	9969	9971	9972	9973	9974	9975	9976	9977	9978
3,7	9978	9980	99	9981	9982	9982	9983	9984	9984	9985
3,8	9986	9986	9987	9987	9988	9988	9989	9989	9990	9990
3,9	9990	9991	9991	9992	9992	9992	9992	9993	9993	9993

Scanned with OKEN Scanner

Tabela V

Valorile χ^2_ε în funcție de probabilitatea

$$P(\chi^2 > \chi^2_\varepsilon) = \frac{1}{\Gamma\left(\frac{k}{2}\right) \cdot 2^{\frac{k}{2}}} \int_{\chi^2_\varepsilon}^{\infty} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} dx.$$

de numărul k al gradelor de libertate ale repartiției χ^2

Nr. gradelor de libertate k	Probabilitatea $P(\chi^2 > \chi^2_\varepsilon)$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30
1	0,00016	0,0006	0,0039	0,016	0,064	0,148	0,455	1,07
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,66
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,2
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,4
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9
12	3,0	4,2	5,2	6,3	7,8	9,0	11,3	14,0
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8
21	8,9	9,9	11,6	13,2	15,4	17,2	20,3	23,9
22	9,5	10,6	12,3	14,0	16,3	18,1	21,3	24,9
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,3
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5
30	15,0	16,3	18,5	20,6	23,4	25,5	29,3	33,5

Scanned with OKEN Scanner

(continuare tabela V)

Nr. gradelor de libertate k	Probabilitatea $P(\chi^2 > \chi^2_\alpha)$							
	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001
1	1,64	2,7	3,8	5,4	6,6	7,9	9,5	10,83
2	3,22	4,6	6,0	7,8	9,2	11,6	12,4	13,8
3	4,64	6,3	7,8	9,8	11,3	12,8	14,8	16,3
4	6,0	7,8	9,5	11,7	13,3	14,9	16,9	18,5
5	7,3	9,2	11,1	13,4	15,1	16,3	18,9	20,5
6	8,6	10,6	12,6	15,0	16,8	18,6	20,7	22,5
7	9,8	12,0	14,1	16,6	18,5	20,3	22,6	24,3
8	11,0	13,4	15,5	18,2	20,1	21,9	24,3	26,1
9	12,2	14,7	16,9	19,7	21,7	23,6	26,1	27,9
10	13,4	16,0	18,3	21,2	23,2	25,2	27,7	29,6
11	14,6	17,3	19,7	22,6	24,7	26,8	29,4	31,3
12	15,8	18,5	21,0	24,1	26,2	28,3	31	32,9
13	17,0	19,8	22,4	25,5	27,7	29,8	32,5	34,5
14	18,2	21,1	23,7	26,9	29,1	31	34	36,1
15	19,3	22,3	25,0	28,3	30,6	32,5	35,5	37,7
16	20,5	23,5	26,3	29,6	32,0	34	37	39,2
17	21,6	24,8	27,6	31,0	33,4	35,5	38,5	40,8
18	22,8	26,0	28,9	32,3	34,8	37	40	42,3
19	23,9	27,2	30,1	33,7	36,2	38,5	41,5	43,8
20	25,0	28,4	31,4	35,0	37,6	40	43	45,3
21	26,2	29,6	32,7	36,3	38,9	41,5	44,5	46,8
22	27,3	30,8	33,9	37,7	40,3	42,5	46	48,3
23	28,4	32,0	35,2	39,0	41,6	44,0	47,5	49,7
24	29,6	33,2	36,4	40,3	43,0	45,5	48,5	51,2
25	30,7	34,4	37,7	41,6	44,3	47	50	52,5
26	31,8	35,6	38,9	42,9	45,6	48	51,5	54,1
27	32,9	36,7	40,1	44,1	47,0	49,5	53	55,5
28	34,0	37,9	41,3	45,6	48,3	51	54,5	56,9
29	35,1	39,1	42,6	46,7	49,6	52,5	56	58,3
30	36,3	40,3	43,8	48,0	50,9	54	57,5	59,7

Scanned with OKEN Scanner

Tabela VI

Probabilitatea $F(t; n)$ pentru repartiția Student, care depinde de t și de numărul n al gradelor de libertate

t \ n	1	2	3	4	5	6	7	8	9	10
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500
0,1	532	535	537	537	538	538	538	539	539	539
0,2	563	570	573	574	575	576	576	577	577	577
0,3	593	604	608	610	612	613	614	614	614	615
0,4	621	636	642	645	647	648	650	650	651	651
0,5	648	667	674	678	681	683	684	685	686	686
0,6	672	695	705	710	713	715	716	717	718	719
0,7	694	722	733	739	742	745	747	748	749	750
0,8	715	746	759	766	770	773	775	777	778	779
0,9	733	768	783	790	795	799	801	803	804	805
1,0	0,750	789	804	813	818	822	825	827	828	830
1,1	765	807	824	834	839	843	846	848	850	851
1,2	779	824	842	852	858	862	865	868	870	871
1,3	791	838	858	868	876	879	883	885	887	889
1,4	803	852	872	883	890	894	898	900	902	904
1,5	813	864	885	896	903	908	911	914	916	918
1,6	822	875	896	908	915	920	923	926	928	930
1,7	831	884	906	918	925	930	934	936	938	940
1,8	839	893	915	927	934	939	943	945	947	949
1,9	846	901	923	935	942	947	950	953	955	957
2,0	0,852	908	930	942	949	954	957	960	962	963
2,2	864	921	942	954	960	965	968	970	972	974
2,4	874	931	952	963	969	973	976	978	980	981
2,6	883	938	960	970	976	980	982	984	986	987
2,8	891	946	966	976	981	984	987	988	990	991
3,0	898	952	971	980	985	988	990	992	992	993
3,2	904	957	975	984	988	991	992	994	995	995
3,4	909	962	979	986	990	993	994	995	996	997
3,6	914	965	982	989	992	994	996	996	997	998
3,8	918	969	984	990	994	996	997	997	998	998
4,0	922	971	986	992	995	996	997	998	998	999
4,2	926	974	988	993	996	997	998	998	999	999
4,4	929	976	989	994	996	998	998	999	999	999
4,6	932	978	990	995	997	998	999	999	999	1,000
4,8	935	980	991	996	998	998	999	999	1,000	
5,0	937	981	992	996	998	999	999	1,003		
5,2	940	982	993	997	998	999	999			
5,4	942	984	994	997	998	999	1 000			
5,6	944	985	994	998	999	999				
5,8	946	986	995	998	999	999				
6,00	947	987	995	998	999	1 000				

Scanned with OKEN Scanner

(continuare tabela VI)

t \ h	11	12	13	14	15	16	17	18	19	
0,0	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,500	0,50000
0,1	539	539	539	539	539	539	539	539	539	53983
0,2	577	578	578	578	578	578	578	578	578	57926
0,3	615	615	616	616	616	616	616	616	616	61791
0,4	652	652	652	652	653	653	653	653	653	65542
0,5	686	687	687	688	688	688	688	688	689	69146
0,6	720	720	721	721	721	722	722	722	722	72575
0,7	751	751	752	752	753	753	753	754	754	75804
0,8	780	780	781	782	782	782	783	783	783	78814
0,9	806	807	808	808	809	809	810	810	810	81594
1,0	831	832	832	833	833	834	834	835	835	84134
1,1	853	854	854	855	856	856	857	857	858	86433
1,2	872	873	874	875	876	876	877	877	878	88493
1,3	890	891	892	893	893	894	894	895	895	90320
1,4	906	907	908	908	909	910	910	911	911	91924
1,5	919	920	921	922	923	924	924	294	925	93319
1,6	931	932	933	934	935	935	936	936	937	94520
1,7	941	943	944	945	945	946	946	947	947	95543
1,8	950	952	952	953	954	955	955	956	956	96407
1,9	958	959	960	961	962	962	963	963	964	97128
2,0	965	967	967	967	968	969	969	970	970	97725
2,2	975	976	977	977	978	979	979	979	980	98610
2,4	982	983	984	985	985	986	986	986	987	99180
2,6	988	988	989	990	990	990	991	991	991	99534
2,8	0,991	0,992	0,992	0,993	0,993	0,994	0,994	0,994	0,994	99744
3,0	994	994	995	995	996	996	996	996	996	0,99865
3,2	996	996	996	997	997	997	997	998	998	99931
3,4	997	997	998	998	998	998	998	998	998	99966
3,6	998	998	998	999	999	999	999	999	999	99984
3,8	998	999	999	999	999	999	999	999	999	99993
4,0	999	999	999	999	999	1,000	1,000	1,000	1,000	99997
4,2	999	999	1,000	1,000	1,000					99999
4,4	1,000	1,000								99999

Scanned with OKEN Scanner

Limitele intervalului de încredere 95% pentru probabilitatea p din distribuția binomială, estimată cu frecvența relativă $\frac{x}{n}$ [1].

$n-x$ \ x	0	1	5	10	20	50	100	200	500	
1	975 000	987 013	996 359	998 587	999 762	1000 896	1000 946	1000 973	1000 989	1000 1000
5	522 000	641 004	813 187	882 384	932 593	970 800	984 892	992 944	997 977	1000 1000
10	308 000	413 002	616 118	728 272	827 472	917 714	955 838	977 914	991 964	1000 1000
15	218 000	302 002	491 087	614 211	737 393	865 649	927 794	961 888	984 953	1000 1000
20	168 000	238 001	407 068	528 173	662 338	816 534	895 755	943 863	976 941	1000 1000
26	132 000	190 001	337 055	452 142	589 289	763 540	860 713	923 836	967 928	1000 1000
30	116 000	167 001	303 048	412 127	548 264	731 510	838 678	910 819	561 920	1000 1000
35	100 000	145 001	268 042	372 112	504 238	694 476	812 658	894 799	954 910	1000 1000
40	088 000	129 001	241 037	338 100	467 217	660 447	787 632	878 780	947 901	1000 1000
45	079 000	115 001	218 033	310 091	434 199	630 421	764 608	863 762	939 891	1000 1000
50	071 000	101 001	200 030	286 083	406 184	602 398	741 585	848 745	932 882	1000 1000
60	060 000	088 000	170 025	248 071	359 160	552 359	700 545	819 713	917 864	1000 1000
80	045 000	067 000	132 019	196 055	292 126	474 301	630 480	766 658	889 831	1000 1000
100	036 000	054 000	108 016	162 045	245 105	415 259	571 429	720 605	862 801	1000 1000
200	018 000	027 000	056 008	086 023	137 057	295 152	395 280	550 450	747 681	1000 1000
500	007 000	011 000	023 003	036 009	059 024	118 068	199 138	319 253	531 469	1000 1000
	000 000	000 000	000 000	000 000	000 000	000 000	000 000	000 000	000 000	— —

[1] Dacă $\frac{x}{n} = \frac{10}{60}$ rezultă $x = 10$, $n = 60$, deci $n - x = 50$; pe coloana 10 și linia 50 găsim $\begin{vmatrix} 917 \\ 714 \end{vmatrix}$, de unde intervalul de încredere 95%: $0{,}714 < p < 0{,}917$.

Tabela VIII

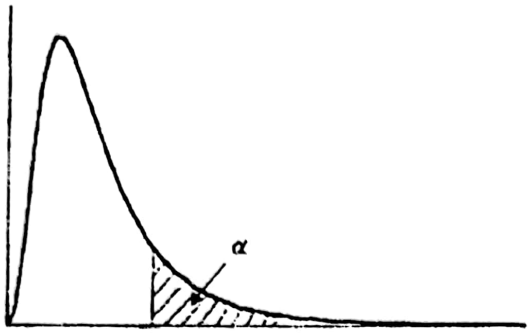
Valorile de 5 % și 1 % ale limitelor abaterii variabilei F.

n/m	1	2	3	4	5	6	7	8	9	10	11	12
1	161 4052	200 4999	216 5403	225 5625	230 5764	234 5859	237 5928	239 5981	241 6022	241 6056	243 6082	244 6106
2	18,51 98,49	19,00 99,01	19,16 99,17	19,25 99,25	19,30 99,30	19,33 99,33	19,36 99,34	19,37 99,36	19,38 99,38	19,39 99,40	19,40 99,41	19,41 99,42
3	10,13 34,12	9,55 30,81	9,28 29,46	9,12 28,71	9,01 28,24	8,94 27,91	8,88 27,67	8,84 27,49	8,81 27,34	8,78 27,23	8,76 27,13	8,74 27,05
4	7,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,26 15,52	6,16 15,21	6,09 14,98	6,04 14,80	6,00 14,66	5,96 14,54	5,93 14,45	5,91 14,37
5	6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	5,05 10,97	4,95 10,67	4,88 10,45	4,82 10,27	4,78 10,15	4,74 10,05	4,70 9,96	4,68 9,89
6	5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,39 8,75	4,28 8,47	4,21 8,26	4,15 8,10	4,10 7,98	4,06 7,87	4,03 7,79	4,00 7,72
7	5,59 12,25	4,74 9,55	4,35 8,45	4,12 7,85	3,97 7,46	3,87 7,19	3,79 7,00	3,73 6,84	3,68 6,71	3,63 6,62	3,60 6,54	3,57 6,47
8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,69 6,63	3,58 6,37	3,50 6,19	3,44 6,03	3,39 5,91	3,34 5,82	3,31 5,74	3,28 5,67
9	5,12 10,56	4,26 8,02	3,86 6,99	3,63 6,42	3,48 6,06	3,37 5,80	3,29 5,62	3,23 5,47	3,18 5,35	3,13 5,26	3,10 5,18	3,07 5,11
10	4,96 10,04	4,10 7,56	3,71 6,55	3,48 5,99	3,33 5,64	3,22 5,39	3,14 5,21	3,07 5,06	3,02 4,95	2,97 4,85	2,94 4,78	2,91 4 71
11	4,84 9,85	3,98 7,20	3,59 6,22	3,36 5,67	3,20 5,32	3,09 5,07	3,01 4,88	2,95 4,74	2,90 4,63	2,86 4,54	2,82 4,46	2,79 4,40
12	4,75 9,33	3,88 6,93	3,49 5,95	3,26 5,41	3,11 5,06	3,00 4,82	2,92 4,65	2,85 4,50	2,80 4,39	2,76 4,30	2,72 4,22	2,69 4,16
13	4,67 9,07	3,80 6,70	3,41 5,74	3,18 5,20	3,02 4,86	2,92 4,62	2,84 4,44	2,77 4,30	2,72 4,19	2,67 4,10	2,63 4,02	2,60 3,96
14	4,60 8,86	3,74 6,51	3,34 5,56	3,11 5,03	2,96 4,69	2,85 4,46	2,77 4,28	2,70 4,14	2,65 4,03	2,60 3,94	2,56 3,86	2,53 3,80
15	4,54 8,68	3,68 6,36	3,29 5,42	3,06 4,89	2,90 4,56	2,79 4,32	2,70 4,14	2,64 4,00	2,59 3,89	2,55 3,80	2,51 3,73	2,48 3,67
16	4,49 8,53	3,63 6,23	3,24 5,29	3,01 4,77	2,85 4,44	2,74 4,20	2,66 4,03	2,59 3,89	2,54 3,78	2,49 3,69	2,45 3,61	2,42 3,55
17	4,45 8,40	3,59 6,11	3,20 5,18	2,96 4,67	2,81 4,34	2,70 4,10	2,62 3,93	2,55 3,76	2,50 3,68	2,45 3,59	2,41 3,52	2,38 3,45
18	4,41 8,28	3,55 6,01	3,16 5,09	2,93 4,58	2,77 4,25	2,66 4,01	2,58 3,85	2,51 3,71	2,46 3,60	2,41 3,51	2,37 3,44	2,34 3,37
19	4,38 8,18	3,52 5,93	3,13 5,01	2,90 4,50	2,74 4,17	2,63 3,94	2,55 3,77	2,48 3,63	2,43 3,52	2,38 3,43	2,34 3,36	2,31 3,30
20	4,35 8,10	3,49 5,85	3,10 4,94	2,87 4,43	2,71 4,10	2,60 3,87	2,52 3,71	2·45 3,56	2,40 3,45	2,35 3,37	2,31 3,30	2,28 3,23
21	4,32 8,02	3,47 5,78	3,07 4,87	2,84 4,37	2,68 4,04	2,57 3,81	2,49 3,65	2,42 3,51	2,37 3,40	2,32 3,31	2,28 3,24	2,25 3,17
22	4,30 7,94	3,44 5,72	3,05 4,82	2,82 4,31	2,66 3,99	2,55 3,76	2,47 3,59	2,40 3,45	2,35 3,35	2·30 3,26	2,26 3,18	2,23 3,12
23	4,28 7,88	3,42 5,66	3,03 4,76	2,80 4,26	2,64 3,94	2,53 3,71	2,45 3,54	2,38 3,41	2,32 3,30	2,28 3,21	2,24 3,14	2,20 3,07
24	4,26 7,82	3,40 5,61	3,01 4,72	2,78 4,22	2,62 3,90	2,51 3,67	2,43 3,50	2,36 3,36	2,30 3,25	2,26 3,17	2,22 3,09	2,18 3,03
25	4,24 7,77	3,38 5,57	2,99 4,68	2,76 4,18	2,60 3,86	2,49 3,63	2,41 3,46	2,34 3,32	2,28 3,21	2,24 3,13	2,20 3,05	2,16 2,99
26	4,22 7,72	3,37 5,53	2,98 4,64	2,74 4,14	2,59 3,82	2,47 3,59	2,39 3,42	2,32 3,29	2,27 3,17	2,22 3,09	2,18 3,02	2,15 2,96

Scanned with OKEN Scanner

în funcție de gradele de libertate m și n. *(continuare Tabela VIII)*

14	16	20	24	30	40	50	75	100	200	500	∞	n/m
245 6142	246 6169	248 6208	249 6234	250 6258	251 6286	252 6302	253 6323	253 6334	254 6352	254 6361	254 6366	1
19,42 99,43	19,43 99,44	19,44 99,45	19,45 99,46	19,46 99,47	19,47 99,48	19,47 99,48	19,48 99,49	19,49 99,49	19,49 99,49	19,50 99,50	19,50 99,50	2
8,71 26,92	8,69 26,83	8,66 26,69	8,64 26,60	8,62 26,50	8,60 26,41	8,58 26,35	8,57 26,27	8,56 26,23	8,54 26,18	8,54 26,14	8,53 26,12	3
5,87 14,24	5,84 14,15	5,80 14,02	5,77 13,93	5,74 13,83	5,71 13,74	5,70 13,69	5,68 13,61	5,66 13,57	5,65 13,52	5,64 13,48	5,63 13,46	4
4,64 9,77	4,60 9,68	4,56 9,55	4,53 9,47	4,50 9,38	4,46 9,29	4,46 9,24	4,42 9,17	4,40 9,13	4,38 9,07	4,37 9,04	4,36 9,02	5
3,96 7,60	3,92 7,52	3,87 7,39	3,84 7,31	3,81 7,23	3,77 7,14	3,75 7,09	3,72 7,02	3,71 6,99	3,69 6,94	3,68 6,90	3,67 6,88	6
3,52 6,35	3,49 6,27	3,44 6,15	3,41 6,07	3,38 5,98	3,34 5,90	3,32 5,85	3,29 5,78	3,28 5,75	3,25 5,70	3,24 5,67	3,23 5,65	7
3,23 5,56	3,20 5,48	3,15 5,36	3,12 5,28	3,08 5,20	3,05 5,11	3,03 5,06	3,00 5,00	2,98 4,96	2,96 4,91	2,96 4,88	2,52 4,86	8
3,02 5,00	2,98 4,92	2,93 4,80	2,90 4,73	2,86 4,64	2,82 4,56	2,80 4,51	2,77 4,45	2,76 4,41	2,73 4,36	2,72 4,33	2,71 4,31	9
2,86 4,60	2,82 4,52	2,77 4,41	2,74 4,33	2,70 4,25	2,67 4,17	2,64 4,12	2,61 4,05	2,59 4,01	2,56 3,96	2,55 39,3	2,54 3,91	10
2,74 4,29	2,70 4,21	2,65 4,10	2,61 4,02	2,57 3,94	2,53 3,86	2,50 3,80	2,47 3,74	2,45 3,70	2,42 3,60	2,41 3,62	2,40 3,60	11
2,64 4,05	2,60 3,98	2,54 3,86	2,50 3,78	2,46 3,70	2,42 3,61	2,40 3,56	2,36 3,49	2,35 3,46	2,32 3,41	2,31 3,38	2,30 3,36	12
2,55 3,85	2,51 3,78	2,46 3,67	2,42 3,59	2,38 3,51	2,34 3,42	2,32 3,37	2,28 3,30	2,26 3,27	2,24 3,21	2,22 3,18	2,21 3,16	13
2,48 3,70	2,44 3,62	2,39 3,51	2,35 3,43	2,31 3,34	2,27 3,26	2,24 3,21	2,21 3,14	2,19 3,11	2,16 3,06	2,14 3,02	2,13 3,00	14
2,43 3,56	2,39 3,48	2,33 3,36	2,29 3,29	2,25 3,20	2,21 3,12	2,18 3,07	2,15 3,00	2,12 2,97	2,10 2,92	2,08 2,89	2,07 2,87	15
2,37 3,45	2,33 3,37	2,28 3,25	2,24 3,18	2,20 3,10	2,16 3,01	2,13 2,96	2,09 2,89	2,07 2,86	2,04 2,80	2,02 2,77	2,01 2,75	16
2,33 3,35	2,29 3,27	2,23 3,16	2,19 3,08	2,15 3,00	2,11 2,92	2,08 2,86	2,04 2,79	2,02 2,76	1,99 2,70	1,97 2,67	1,96 2,65	17
2,29 3,27	2,25 3,19	2,19 3,07	2,15 3,00	2,11 2,91	2,07 2,80	2,04 2,75	2,00 2,71	1,98 2,68	1,95 2,62	1,93 2,59	1,92 2,57	18
2,26 3,19	2,21 3,12	2,15 3,00	2,11 2,92	2,07 2,84	2,02 2,76	2,00 2,70	1,96 2,63	1,94 2,60	1,91 2,54	1,90 2,51	1,88 2,49	19
2,23 3,13	2,18 3,05	2,12 2,94	2,08 2,86	2,04 2,77	1,99 2,69	1,96 2,63	1,92 2,56	1,90 2,53	1,87 2,47	1,85 2,44	1,84 2,42	20
2,20 3,07	2,15 2,99	2,09 2,88	2,05 2,80	2,00 2,72	1,96 2,63	1,93 2,58	1,89 2,51	1,87 2,47	1,84 2,42	1,82 2,38	1,81 2,36	21
2,18 3,02	2,13 2,94	2,07 2,83	2,03 2,75	1,98 2,67	1,93 2,58	1,91 2,53	1,87 2,46	1,84 2,42	1,81 2,37	1,80 2,33	1,78 2,31	22
2,14 2,97	2,10 2,89	2,04 2,78	2,00 2,70	1,96 2,62	1,91 2,53	1,88 2,48	1,84 2,41	1,82 2,37	1,79 2,32	1,77 3,28	1,76 2,26	23
2,13 2,93	2,09 2,86	2,02 2,74	1,98 2,66	1,94 2,58	1,89 2,49	1,86 2,44	1,82 2,36	1,80 2,33	1,76 2,27	1,74 2,23	1,73 2,21	24
2,11 2,89	2,06 2,81	2,00 2,70	1,96 2,62	1,92 2,54	1,87 2,45	1,84 2,40	1,80 2,32	1,77 2,29	1,74 2,13	1,72 2,19	1,71 2,17	25
2,10 2,86	2,05 2,74	2,09 2,61	1,95 2,58	1,90 2,50	1,85 2,41	1,82 2,36	1,78 2,28	1,76 2,25	1,72 2,19	1,70 2,15	1,69 2,13	26

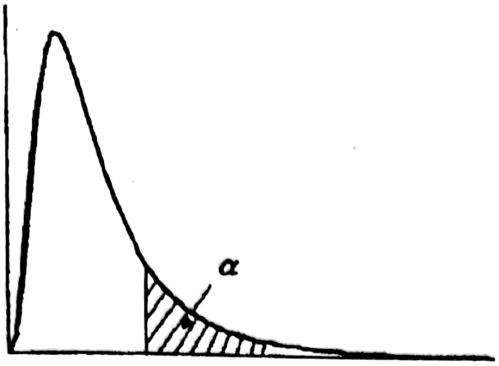


(Continuare Tabela VIII)

Valorile distribuției F, $\alpha = 0,1$

n/m	1	2	3	4	5	6	7	8	9
1	39.864	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.857
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965
21	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948
22	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933
23	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895
26	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884
27	2.901	2.511	2.299	2.165	2.073	2.005	1.952	1.909	1.874
28	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865
29	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738
120	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684
∞	2.706	2.303	2.084	1.945	1.847	1.774	1.717	1.670	1.632

Scanned with OKEN Scanner



(Continuare Tabela VIII)

Valorile distribuției F, $\alpha = 0{,}025$

n/m	1	2	3	4	5	6	7	8	9
1	647.793	799.482	864.151	899.599	921.835	937.114	948.203	956.643	963.279
2	38.506	39.000	39.166	39.248	39.298	39.331	39.356	39.373	39.387
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473
4	12.218	10.649	9.979	9.604	9.364	9.197	9.074	8.980	8.905
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929
19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837
21	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798
22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763
23	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677
26	5.659	4.265	3.670	3.329	3.105	2.945	2.824	2.729	2.653
27	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631
28	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611
29	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.669	2.592
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334
120	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222
∞	5.024	3.689	3.116	2.786	2.567	2.408	2.288	2.192	2.114

Scanned with OKEN Scanner

(Continuare Tabela VIII)

Valorile distribuției F, $\alpha = 0{,}025$

n/m	10	12	15	20	24	30	40	60	120	∞
1	60.195	60.705	61.220	61.740	62.002	62.265	62.529	62.794	63.061	63.328
2	9.392	9.408	9.425	9.441	9.450	9.458	9.466	9.475	9.483	9.491
3	5.230	5.216	5.200	5.184	5.176	5.168	5.160	5.151	5.143	5.134
4	3.920	3.896	3.870	3.844	3.831	3.817	3.804	3.790	3.775	3.761
5	3.297	3.268	3.238	3.207	3.191	3.174	3.157	3.140	3.123	3.105
6	2.937	2.905	2.871	2.836	2.818	2.800	2.781	2.762	2.742	2.722
7	2.703	2.668	2.632	2.595	2.575	2.555	2.535	2.514	2.493	2.471
8	2.538	2.502	2.464	2.425	2.404	2.383	2.361	2.339	2.316	2.293
9	2.416	2.379	2.340	2.298	2.277	2.255	2.232	2.208	2.184	2.159
10	2.323	2.284	2.244	2.201	2.178	2.155	2.132	2.107	2.082	2.055
11	2.248	2.209	2.167	2.123	2.100	2.076	2.052	2.026	2.000	1.972
12	2.188	2.147	2.105	2.060	2.036	2.011	1.986	1.960	1.932	1.904
13	2.138	2.097	2.053	2.007	1.983	1.958	1.931	1.904	1.876	1.846
14	2.095	2.054	2.010	1.962	1.938	1.912	1.885	1.857	1.828	1.797
15	2.059	2.017	1.972	1.924	1.899	1.873	1.845	1.817	1.787	1.755
16	2.028	1.985	1.940	1.891	1.866	1.839	1.811	1.782	1.751	1.718
17	2.001	1.958	1.912	1.862	1.836	1.809	1.781	1.751	1.719	1.686
18	1.977	1.933	1.887	1.837	1.810	1.783	1.754	1.723	1.691	1.657
19	1.956	1.912	1.865	1.814	1.787	1.759	1.730	1.699	1.666	1.631
20	1.937	1.892	1.845	1.794	1.767	1.738	1.708	1.677	1.643	1.607
21	1.920	1.875	1.827	1.776	1.748	1.719	1.689	1.657	1.623	1.586
22	1.904	1.859	1.811	1.759	1.731	1.702	1.671	1.639	1.604	1.567
23	1.890	1.845	1.796	1.744	1.716	1.686	1.655	1.622	1.587	1.549
24	1.877	1.832	1.783	1.730	1.702	1.672	1.641	1.607	1.571	1.533
25	1.866	1.820	1.771	1.718	1.689	1.659	1.627	1.593	1.557	1.518
26	1.855	1.809	1.760	1.706	1.677	1.647	1.615	1.581	1.544	1.504
27	1.845	1.799	1.749	1.695	1.666	1.636	1.603	1.569	1.531	1.491
28	1.836	1.790	1.740	1.685	1.656	1.625	1.592	1.558	1.520	1.478
29	1.827	1.781	1.731	1.676	1.647	1.616	1.583	1.547	1.509	1.467
30	1.819	1.773	1.722	1.667	1.638	1.606	1.573	1.538	1.499	1.456
40	1.763	1.715	1.662	1.605	1.574	1.541	1.506	1.467	1.425	1.377
60	1.707	1.657	1.603	1.543	1.511	1.476	1.437	1.395	1.348	1.291
120	1.652	1.601	1.545	1.482	1.447	1.409	1.368	1.320	1.265	1.193
∞	1.599	1.546	1.487	1.421	1.383	1.342	1.295	1.240	1.169	1.000

Scanned with OKEN Scanner

Probabilitățile $P_n(\lambda) = K(\lambda)$ ale cepartiției lui Kolmogorov

λ	$K(\lambda)$	λ	$K(\lambda)$	λ	$K(\lambda)$
0,28	0,000001	0,71	0,305471	1,14	0,861394
0,29	0,000004	0,72	0,322265	1,15	0,868038
0,30	0,000009	0,73	0,339118	1,16	0,864442
0,31	0,000021	0,74	0,355981	1,17	0,870612
0,32	0,000046	0,75	0,372833	1,18	0,876548
0,33	0,000091	0,76	0,389640	1,19	0,882258
0,34	0,000171	0,77	0,406372	1,20	0,887750
0,35	0,000303	8,78	0,423002	1,21	0,893030
0,36	0,000511	0,79	0,439505	1,22	0,898104
0,37	0,000826	0,80	0,455857	1,23	0,902972
0,38	0,001285	0,81	0,472041	1,24	0,907648
0,39	0,001929	0,82	0,488030	1,25	0,912132
0,40	0,002808	0,83	0,503808	1,26	0,916432
0,41	0,003972	0,84	0,519366	1,27	0,920556
0,42	0,005476	0,85	0,534682	1,28	0,924505
0,43	0,007377	0,86	0,549744	1,29	0,928288
0,44	0,009430	0,87	0,564546	1,30	0,931908
0,45	0,012590	0,88	0,579070	1,31	0,935370
0,46	0,016005	0,89	0,593316	1,32	0,938682
0,47	0,020022	0,90	0,607270	1,33	0,941848
0,48	0,024682	0,91	0,620928	1,34	0,944872
0,49	0,030017	0,92	0,634286	1,35	0,947756
0,50	0,036055	0,93	0,647338	1,36	0,950512
0,51	0,042814	0,94	0,660082	1,37	0,953142
0,52	0,050306	0,95	0,672516	1,38	0,955650
0,53	0,058534	0,96	0,684636	1,39	0,958040
0,54	0,067497	0,97	0,696444	1,40	0,960318
0,55	0,077183	0,98	0,707940	1,41	0,962486
0,56	0,087577	0,99	0,719126	1,42	0,964552
0,57	0,098656	1,00	0,730000	1,43	0,966516
0,58	0,110395	1,01	0,740566	1,44	0,968382
0,59	0,122760	1,02	0,750826	1,45	0,970158
0,60	0,135718	1,03	0,760780	1,46	0,971846
0,61	0,149229	1,04	0,770434	1,47	0,973448
0,62	0,163225	1,05	0,779794	1,48	0,974970
0,63	0,177753	1,06	0,788860	1,49	0,976412
0,64	0,192677	1,07	0,797636	1,50	0,977782
0,65	0,207987	1,08	0,806128	1,51	0,979080
0,66	0,223637	1,09	0,814342	1,52	0,980310
0,67	0,239582	1,10	0,822282	1,53	0,981476
0,68	0,255780	1,11	0,829950	1,54	0,982578
0,69	0,272189	1,12	0,837356	1,55	0,983622
0,70	0,288765	1,13	0,844502	1,56	0,984610

Scanned with OKEN Scanner

(continuare tabela IX)

λ	K(λ)	γ	K(λ)	γ	K(π)
1,57	0,985544	1,92	0,998745	2,27	0,999934
1,58	0,986426	1,93	0,998837	2,28	0,999940
1,59	0,987260	1,94	0,998924	2,29	0,999944
1,60	0,988048	1,95	0,999001	2,30	0,999949
1,61	0,988791	1,96	0,999079	2,31	0,999954
1,62	0,989492	1,97	0,999149	2,32	0,999958
1,63	0,990154	1,98	0,999213	2,33	0,999962
1,64	0,990777	1,99	0,999273	2,34	0,999965
1,65	0,991364	2,00	0,999329	2,35	0,999968
1,66	0,991917	2,01	0,999380	2,36	0,999970
1,67	0,992438	2,02	0,999428	2,37	0,999973
1,68	0,992928	2,03	0,999474	2,38	0,999976
1,69	0,993389	2,04	0,999516	2,39	0,999978
1,70	0,993823	2,05	0,999552	2,40	0,999980
1,71	0,994230	2,06	0,999588	2,41	0,999982
1,72	0,994612	2,07	0,999620	2,42	0,999984
1,73	0,994972	2,08	0,999650	2,43	0,999986
1,74	0,995309	2,09	0,999680	2,44	0,999987
1,75	0,995625	2,10	0,999705	2,45	0,999988
1,76	0,995922	2,11	0,999728	2,46	0,999988
1,77	0,996200	2,12	0,999750	2,47	0,999990
1,78	0,996460	2,13	0,999770	2,48	0,999991
1,79	0,996704	2,14	0,999790	2,49	0,999992
1,80	0,996932	2,15	0,999806	2,50	0,9999925
1,81	0,997146	2,16	0,999822	2,55	0,9999956
1,82	0,997346	2,17	0,999838	2,60	0,9999974
1,83	0,997533	2,18	0,999852	2,65	0,9999984
1,84	0,997707	2,19	0,999864	2,70	0,9999990
1,85	0,997870	2,20	0,999874	2,75	0,9999994
1,86	0,998023	2,21	0,999886	2,80	0,9999997
1,87	0,998145	2,22	0,999896	2,85	0,99999982
1,88	0,998297	2,23	0,999904	2,90	0,99999990
1,89	0,998421	2,24	0,999912	2,95	0,99999994
1,90	0,998536	2,25	0,999920	3,00	0,99999997
1,91	0,998644	2,26	0,999926		

Scanned with OKEN Scanner

BIBLIOGRAFIE

1. Andrei, T., Stancu, S., *Statistică — teorie şi aplicaţii*, Ed. All, Bucureşti, 1995
2. Anghelache, C., Niculescu, E., *Breviar statistic*, Ed. Economică, Bucureşti, 2000
3. Baron, T., Biji, E., Tövissi, L., ş.a., *Statistică teoretică şi economică*, E.D.P. Bucureşti, 1991
4. Bădiţă, M., Baron, T., Korka, M., *Statistică pentru afaceri*, Ed. Eficient, Bucureşti, 1998
5. Capanu, I., Wagner, P., Secăreanu, C., *Statistică macroeconomică*, Ed. Economică, Bucureşti, 1997
6. Ciucu, G., Craiu, V., Ştefănescu, V., Statistică matematică şi cercetări operaţionale, E.D.P., 1978
7. Craiu, V., *Verificarea ipotezelor statistice,* E.D.P., Bucureşti, 1972
8. Dodge, Y., *Statistique, Neuchâtel,* P.A.N., 1990
9. Faverge, J., M., *Méthodes statistiques en psychologie appliquée,* Paris, P.U.F., 1963
10. Fisz, M., *Probability and mathematical statistics*, New-York, 1964
11. Goleţ, I., *Statistică şi Probabilităţi Teorie şi aplicaţii în economie,* Editura Politehnica, colecţia „STUDENT", 2001
12. Grosbras, I.M., *Méthodes statistiques des sondages*, Ed Economică, Paris, 1987
13. Guilford, J. P., Fruchter, B., *Fundamental statistics in psychology and education*, MC. Graw-Hill, 1973
14. Ionescu, H., M., *Statistică matematică*, E.D.P., Bucureşti, 1962
15. Jaba, E., *Statistica*, Editura Economică, Bucureşti, 1998
16. Mihoc, Gh., Ciucu, G., Craiu, V., *Teoria probabilităţilor şi statistica matematică,* E.D.P., Bucureşti, 1970
17. Mihoc, Gh., Urseanu, V., *Matematici aplicate în statistică*, Ed. Academiei Române, Bucureşti, 1962

Scanned with OKEN Scanner

CUPRINS

Prefață / 5
CAPITOLUL 1. INTRODUCERE ÎN STATISTICĂ / 7
1.1. Chestiuni preliminarii / 7
1.2. Istoricul statisticii / 9
1.3. Obiectul şi metoda statisticii / 16
1.4. Categoriile ştiinţei statistice / 19
CAPITOLUL 2. SERII STATISTICE. INDICATORI NUMERICI / 24
2.1. Serii statistice / 25
2.2. Indicatori numerici (parametri) pentru serii statistice cu o singură caracteristică / 31
2.3. Piramida vârstelor / 49
2.4. Probleme / 53
CAPITOLUL 3. ELEMENTE DE TEORIA PROBABILITĂȚILOR. VARIABILE ALEATOARE / 62
3.1. Teoria probabilităților / 62
3.2. Variabile aleatoare. Noțiuni generale / 66
3.3. Distribuții (Repartiții) discrete şi continue. Proprietăți / 77
3.4. Programe / 92
3.4.1. Comparație între histograma seriei şi distribuția normală / 92
3.4.2. Repartiția normală / 94
3.4.3. Repartiția exponențială / 95
3.4.4. Repartiția student (t) / 96
3.4.5. Repartiția hi-pătrat (x^2) / 97
3.4.6. Repartiția F / 98
CAPITOLUL 4. SELECȚIE. TEORIA ELEMENTARĂ A SONDAJULUI / 100
4.1. Noțiuni generale. Indicatori / 100
4.2. Teoria elementară a sondajului / 111
4.3. Aplicații / 134
4.4. Probleme. Teoria elementară a sondajului / 142

Scanned with OKEN Scanner

CAPITOLUL 5. ESTIMAȚII / 148
5.1. Formularea problemei estimației / 148
5.2. Tipuri de estimatori / 150
5.3. Estimare punctuală / 154
5.4. Metoda verosimilității maxime / 155
CAPITOLUL 6. IPOTEZE STATISTICE / 161
6.1. Noțiuni generale / 161
6.2. Erori și riscuri / 164
6.3. Puterea unui test. Cel mai puternic test / 164
6.4. Ipoteze asupra parametrilor repartiției normale / 168
6.5. Teste de semnificație / 176
CAPITOLUL 7. REPARTIȚII ȘI TEOREME LIMITĂ (TL) / 193
7.1. Considerații generale / 193
7.2. Proprietățile distribuției binomiale și distribuției normale / 196
7.3. Aplicații / 200
CAPITOLUL 8. ANALIZA DE REGRESIE ȘI CORELAȚIE / 204
8.1. Probleme ale analizei de regresie și corelație / 204
8.1.1. Noțiuni de bază / 205
8.1.2. Metode analitice de analiză a legăturilor statistice / 206
8.2. Modele de regresie cu două variabile / 208
8.2.1.Modelul de regresie liniară / 208
8.2.2. Modele de regresie neliniară / 210
8.3. Măsurarea intensității legăturii dintre două variabile / 211
8.3.1. Coeficientul de corelație / 211
8.3.2. Raportul de corelație / 212
8.4. Corelația neparametrică / 213
8.5. Corelație și regresie (extensii) / 218
CAPITOLUL 9. SERII CRONOLOGICE / 234
9.1. Indicatori pentru caracterizarea nivelului și variației în timp / 234
9.1.1. Indicatori absoluți / 235
9.1.2. Indicatori relativi / 236
9.1.3. Indicatori medii / 237
9.2. Indicii dinamicii / 245

Scanned with OKEN Scanner

9.2.1. Categorii de indici / 246
9.2.2. Indici agregați / 246
9.2.3. Indici calculați ca mărimi medii / 247
9.2.4. Indici calculați ca raport de medii / 248
9.2.5. Indici după felul bazei de calcul / 249
CAPITOLUL 10. ANALIZA ŞI EXTRAPOLAREA SERIILOR CRONOLOGICE / 250
10.1. Tipologia seriilor cronologice / 250
10.1.1. Componentele unei serii cronologice / 250
10.1.2. Componenta tendențială (trend) / 251
10.1.3. Componenta ciclică / 251
10.1.4. Componenta sezonieră / 252
10.1.5. Componenta aleatoare (reziduală) / 253
10.2. Ajustarea seriilor cronologice / 255
10.2.1. Metoda grafică de ajustare / 255
10.2.2. Metoda analitică de ajustare / 256
CAPITOLUL 11. ANALIZA DISPERSIONALĂ / 262
11.1. Noțiuni generale / 263
11.2. Analiza dispersională unifactorială / 266
11.3. Analiza dispersională bifactorială / 273
11.4. Aplicații / 278
CAPITOLUL 12. ANALIZA CLUSTER / 287
12.1. Noțiuni generale / 287
12.2. Program Cluster / 294
CAPITOLUL 13. BALANȚA DE PLĂȚI EXTERNE ANALIZĂ STATISTICĂ / 299
13.1. Sistemul conceptual şi tipologia BPE / 299
13.2. Structura unei B.P.E / 301
13.3. Echilibrarea balanței globale de P.E. / 305
13.4. Analiza statistică a BPE / 306
13.5. Program pentru analiza statistică a B.P.E. a României / 311
ANEXA 1. UTILIZAREA MATHCAD ÎN STATISTICĂ / 315
ANEXA 2. SPSS / 321
ANEXA 3 ERORI / 324
ANEXA 4.. PROBABILITĂȚI ŞI REPARTIȚII TABELATE / 338
BIBLIOGRAFIE / 355
CUPRINS / 357

Scanned with OKEN Scanner

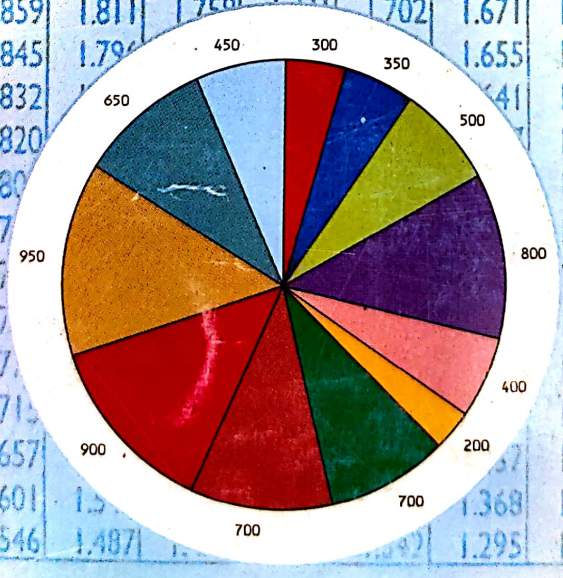
Consilier editorial: ADRIAN DINU RACHIERU
Tehnoredactare: MARIANA LEŞE
Tipografia: **VATRAN ALLIANZ S.R.L.**
1900 Timişoara, str. Axente Sever nr. 9 A
Telefon: 19 06 64; tel/fax.: 19 06 63

360

Scanned with OKEN Scanner

Valorile distribuției F, $\alpha = 0{,}025$

n/m	10	12	15	20	24	30	40	60	120	∞
1	60.195	60.705	61.220	61.740	62.002	62.265	62.529	62.794	63.061	63.328
2	9.392	9.408	9.425	9.441	9.450	9.458	9.466	9.475	9.483	9.491
3	5.230	5.216	5.200	5.184	5.176	5.168	5.160	5.151	5.143	5.134
4	3.920	3.896	3.870	3.844	3.831	3.817	3.804	3.790	3.775	3.761
5	3.297	3.268	3.238	3.207	3.191	3.174	3.157	3.140	3.123	3.105
6	2.937	2.905	2.871	2.836	2.818	2.800	2.781	2.762	2.742	2.722
7	2.703	2.668	2.632	2.595	2.575	2.555	2.535	2.514	2.493	2.471
8	2.538	2.502	2.464	2.425	2.404	2.383	2.361	2.339	2.316	2.293
9	2.416	2.379	2.340	2.298	2.277	2.255	2.232	2.208	2.184	2.159
10	2.323	2.284	2.244	2.201	2.178	2.155	2.132	2.107	2.082	2.055
11	2.248	2.209	2.167	2.123	2.100	2.076	2.052	2.026	2.000	1.972
12	2.188	2.147	2.105	2.060	2.036	2.011	1.986	1.960	1.932	1.904
13	2.138	2.097	2.053	2.007	1.983	1.958	1.931	1.904	1.876	1.846
14	2.095	2.054	2.010	1.962	1.938	1.912	1.885	1.857	1.828	1.797
15	2.059	2.017	1.972	1.924	1.899	1.873	1.845	1.817	1.787	1.755
16	2.028	1.985	1.940	1.891	1.866	1.839	1.811	1.782	1.751	1.718
17	2.001	1.958	1.912	1.862	1.836	1.809	1.781	1.751	1.719	1.686
18	1.977	1.933	1.887	1.837	1.810	1.783	1.754	1.723	1.691	1.657
19	1.956	1.912	1.865	1.814	1.787	1.759	1.730	1.699	1.666	1.631
20	1.937	1.892	1.845	1.794	1.767	1.738	1.708	1.677	1.643	1.607
21	1.920	1.875	1.827	1.776	1.748	1.719	1.689	1.657	1.623	1.586
22	1.904	1.859	1.811	1.759	1.731	1.702	1.671	1.639	1.604	1.567
23	1.890	1.845	1.794	1.742	1.714	1.686	1.655	1.622	1.587	1.549
24	1.877	1.832	1.781	1.729	1.700	1.672	1.641	1.607	1.571	1.533
25	1.866	1.820	1.769	1.717	1.687	1.659	1.627	1.593	1.557	1.518
26	1.855	1.809	1.758	1.705	1.676	1.647	1.615	1.581	1.544	1.504
27	1.845	1.797	1.746	1.694	1.665	1.636	1.603	1.569	1.531	1.491
28	1.836	1.787	1.735	1.683	1.654	1.624	1.592	1.558	1.520	1.478
29	1.827	1.778	1.726	1.673	1.643	1.614	1.581	1.547	1.509	1.467
30	1.819	1.770	1.718	1.664	1.634	1.604	1.571	1.538	1.499	1.456
40	1.763	1.715	1.661	1.606	1.575	1.544	1.510	1.467	1.425	1.377
60	1.707	1.657	1.603	1.547	1.515	1.482	1.447	1.395	1.348	1.291
120	1.652	1.601	1.545	1.488	1.455	1.421	1.384	1.320	1.265	1.193
∞	1.599	1.546	1.487	1.428	1.394	1.358	1.320	1.295	1.240	1.169	1.000



Editura AVGVSTA Timișoara
2001
ISBN 973-8284-60-6

Scanned with OKEN Scanner